Structural and functional characterization of a novel regulator of linear ubiquitin chain formation

Wei Song

Supervisor: Dr. Benjamin Stieglitz

Submitted in partial fulfilment of the requirements of the Degree of Doctor of Philosophy

School of Biological and Chemical Sciences,
Queen Mary, University of London

September 2020

Statement of originality

I, Wei Song, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature:

Date: 9th March 2021

Details of collaboration and publications:

Kliza, K., Song, W., Pinzuti, I., Schaubeck, S., Kunzelmann, S., Kuntin, D., Fornili, A., Pandini, A., Garnett, J., Hofmann, K., Stieglitz, B., & Husnjak, K. N4BP1 is a unique, dimerization-dependent linear ubiquitin reader that regulates TNFR1 signalling through linear ubiquitin binding and Caspase-8-dependent processing (Manuscript in preparation).

Abstract

LUBAC (Linear Ubiquitin Assembly Complex) is a ubiquitin E3 ligase complex with crucial roles in the initiation of the innate immune response and the activation of the NF-κB signaling pathway. LUBAC displays the unique feature to generate the M1-linked ubiquitin chains (also called the linear ubiquitin chains).

N4BP1 (NEDD4 binding protein 1) is identified as a ubiquitin-binding protein and acts as a novel regulator to inhibits LUBAC. However, the functional role of N4BP1 in the context of its ubiquitin-binding properties remains unclear. In this thesis, biochemical, biophysical and structural approaches were employed to study the function of N4BP1.

This study has shown that the N-terminal part of N4BP1 is required for inhibition of the E3 ligase activity of LUBAC. The KH-like domain displays weak inhibition on the LUBAC core catalytic subunit HOIP. However, it is the UBA-like domain that is defined as the main inhibitory module. The inhibitory effect is modulated by the binding interface regions of N4BP1 and HOIP. All three domains of the N-terminal part of N4BP1 work synergistically together to mediate the inhibitory effect on HOIP.

The C-terminus of N4BP1 which contains both the RNAse domain and the CUE domain, recognizes linear ubiquitin chains with high specificity. Quantitative binding studies have demonstrated that the CUE domain differentiates K48-linked di-ubiquitin chains from mono-ubiquitin, M1- and K63-linked di-ubiquitin chains. The structure of the CUE domain was determined by solution NMR. Key interface residues involved in ubiquitin binding were identified by NMR titrations, which allowed the establishment of a docking model of the CUE domain in complex with mono-ubiquitin. The model highlights the importance of the FP (phenylalanine-proline) motif of

the CUE domain and the hydrophobic patch which surrounds isoleucine 44 of ubiquitin.

Importantly, it has been shown that the RNAse domain of N4BP1 is required to create a binding preference for M1-diUb. The RNAse domain itself does not interact with ubiquitin. Instead, the RNAse domain mediates oligomerization of N4BP1. This self-association creates a spatial arrangement of the CUE domains which is optimized for recognizing linear ubiquitin chains with high affinity.

These findings explain how N4BP1 facilitates its regulatory role in the TNFR signaling pathway and expand our knowledge about the functions of N4BP1 in inflammation and innate immunity.

Acknowledgement

First of all, I'd like to show my massive thanks to my supervisor, Dr. Ben Stieglitz, who is kind enough to accept me as a PhD student in his wonderful research group when I decided to change my supervision. I am very lucky to work with ubiquitin again, which made my PhD life much easier to continue.

My sincere thanks to my dear colleague Dr. Irene Pinzuti. Without her kind guidance, I wouldn't be able to join Ben's group. She is always there generously enough to help me either in the lab or in life. I will never forget our friendship.

My huge thanks to Dr. James Garnetts from King's College London and Dr. Alain Oregioni from the Francis Crick Institute, who have both kindly and patiently instructed me with NMR data collecting, processing and analyzing. Lots of thanks to Dr. Ruth Rose and Dr. Petra Ungerer for the daily management of the lab and providing such a nice environment for us. Thanks to Dr. Simone Kunzelmann from the Francis Crick Institute for the kind help with SPR. Thanks to Dr. Arianna Fornili from QMUL for helping with the structure modelling and thanks to Theo Portlock from Garnetts's group for helping with NMR. My big thanks to all the members in Ben's group. I really enjoying working with everyone in the lab.

I'd also like to thank my ex-supervisor Maxie Roessler. Unfortunately, I didn't manage to stay in her group for the rest of my PhD. However, working on bioenergetics was very attractive and had broadened my knowledge in addition to plant biology and structural biology. Most importantly of all, without her offering me this PhD opportunity, I wouldn't be here studying in QMUL.

To my mom and dad, you are always being there supporting me.

Finally, I cannot image how I survived from this crazy PhD life. I thank to myself, for the determinations to follow the heart, to find the best of me.

Table of contents

Abbreviations

ADMA	Asymmetric dimethylarginine
AdoMet	S-adenosylmethionine
AIP4	Atrophin-1-interacting protein 4
AIR	Ambiguous interaction restraint
Akt	RAC-alpha serine/threonine-protein kinase
AMFR (or gp78 or RNF45)	Autocrine motility factor receptor
AMP	Adenosine mono-phosphate
AMPK	AMP-activated protein kinase
AMSH-LP	AMSH-like protease
APC/C	Anaphase-promoting complex (or cyclosome)
Atg8	Autophagy gene 8
ATP	Adenosine triphosphate
Axin	Axis inhibition protein 1
BARD1	BRCA1-associated RING domain protein 1
BRCA1(or RNF53)	Breast cancer type 1 susceptibility protein
CAS9	CRISPR associated protein 9
CASP8	Caspase 8
Cbl-b	Casitas B-lineage lymphoma proto-oncogene b
CD40-L	CD40 ligand
CDC34	Cell division control protein 34
CDS	Coding sequence
CEZANNE	Cellular zinc finger anti-NF-kappa-B protein
cGAS	Cyclic GMP-AMP synthase
cGAMP	Cyclic guanosine monophosphate–adenosine monophosphate
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
cIAP1/2	Cellular inhibitor of apoptosis protein-1 and 2
CoCUN	Cousin of CUBAN
CoQ	Coenzyme Q
CoZi	Coil-zipper domain
CRISPR	Clustered regularly interspaced short palindromic repeats
CRL	Cullin/RING ubiquitin ligase
Crn7	Coronin 7
CSP	Chemical shift perturbation
CTD	C-terminal domain
CUBAN	Cullin-binding domain associating with NEDD8
CUE	Coupling of ubiquitin to ER degradation domain
Cue1p (or Cue1)	Coupling of ubiquitin conjugation to ER degradation protein 1
Cue2p (or Cue2)	Coupling of ubiquitin conjugation to ER degradation protein 2
CUL	Cullin
CV	Column volume
CW-EPR	Continuous wave EPR

Da	Dalton
DDM	N-dodecyl-β-D-maltoside
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSB	Double-strand break
Dsk2	Ubiquitin domain-containing protein DSK2
DTT	Dithiothreitol
DTX	Protein deltex-1
DUB	Deubiquitinating enzyme
E6AP (or UBE3A)	Ubiquitin protein ligase E3A
E.coli	*Escherichia coli*
EDTA	Ethylene Diamine Tetraacetic Acid
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
E_m	Midpoint potential
$E_{m,7}$	Midpoint potential at pH7
EMBOSS	European molecular biology open software suite
EPR	Electron paramagnetic resonance
ESR	Electron spin resonance
ETC	Electron transport chain
EuDTPA	Europium-diethylenetriaminepentaacetic acid
FAD	Flavin adenine dinucleotide
FADH2	Reduced form of flavin adenine dinucleotide
FT	Fourier transform
GF	Gel filtration
gp78	Glycoprotein 78
GSH	Glutathione
GST	Glutathione-S-transferase
H2A	Histone H2A
H2AX	H2A histone family member X
H2B	Histone H2B
H3K4	Histone H3 lysine K4
H3K79	Histone H3 lysine K79
HCCH-TOCSY	^{1}H-^{13}C-^{13}C-^{1}H total correlation spectroscopy
HECT	Homologous to E6-AP carboxyl terminus
HEK293T	Human embryonic kidney 293 cell
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HHARI (or ARIH1)	Ariadne RBR E3 Ubiquitin Protein Ligase 1
HHR23A (or RAD23A)	UV excision repair protein RAD23 homolog A
HIV-1	Human immunodeficiency virus-1
Mw	Molecular weight
HMW	High molecular weight
HOIL-1L	Heme-oxidized IRP2 (iron regulatory protein 2) ligase 1
HOIP	HOIL-1L interacting protein

hPRC1L	human Polycomb repressive complex 1-like
HRS (or HGS)	Hepatocyte growth factor-regulated tyrosine kinase substrate
HRV	Human rhinovirus
HSQC	Heteronuclear single quantum correlation
HSV-1	Herpes simplex virus 1
HUWE1	HECT, UBA and WWE domain containing E3 ubiquitin protein ligase 1
IBR	In-between-RING domain
IEX	Ion-exchange chromatography
IFN	Interferon
IFN-β	Interferon-β
IKK	IκB kinase complex
IL-1β	Interleukin-1β
IL-6	Interleukin-6
IL-12β	Interleukin-12β
INSIG1	insulin-induced gene 1
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRF-3	Interferon regulatory factor 3
ISG15	Interferon stimulated gene 15
IsoT	Isopeptidase T
ITC	Isothermal titration calorimetry
ITCH	E3 ubiquitin-protein ligase Itchy homolog
IκB	Inhibitor of κB
JNK	c-Jun N-terminal kinase
K48-diUb	K48-linked di-ubiquitin
K63-diUb	K63-linked di-ubiquitin
Kd	Dissociation constant
K_D	Equilibrium dissociation constant (used for SPR only)
KDa	Kilo-Dalton
KH	K Homology domain
KHNYN	KH and NYN domain-containing protein
KLHL20	Kelch-like protein 20
LB	Luria-Bertani broth
LDD	Linear ubiquitin chain determining domain
LKB1	Liver kinase B1 homolog
LMB	The MRC laboratory of molecular biology
LPS	Lipopolysaccharide
LRP5	Low-density lipoprotein receptor-related protein 5
LRP6	Low-density lipoprotein receptor-related protein 6
LTM	LUBAC-tethering motif
LUBAC	Linear ubiquitin chain assembly complex
LUBID	Linear ubiquitin binding domain
M1-diUb	M1-linked di-ubiquitin
MALLS	Multi-angle laser light scattering
MALT1	Mucosa-associated lymphoid tissue lymphoma translocation protein 1

MCPIP	MCP-1-induced protein
MDM2	Mouse double minute 2 homolog
MEF	Mouse Primary Embryonic Fibroblast
MES	2-(N-morpholino)ethanesulfonic acid
Met4	Methionine-requiring protein 4
Mfn2	Mitofusion-2
MidA	Mitochondrial dysfunction gene A
mM	Milli-molar
MMA	Mono-methylarginine
monoUb	mono-ubiquitin
MOPS	3-(*N*-morpholino)propanesulfonic acid
MRI	Magnetic resonance imaging
Mw	Molecular weight
N4BP1	Nedd4-binding protein 1
NAD	Nicotinamide-adenine dinucleotid
NADH	Reduced form of nicotinamide-adenine dinucleotid
NCD	N-terminal conserved domain
NDUFAF7	NADH dehydrogenase [ubiquinone] complex I, assembly factor 7
Nedd4	Neural precursor cell expressed developmentally down-regulated protein 4
NEDD8	Neural precursor cell-expressed developmentally downregulated protein 8
NEMO	NF-κB essential modulator
NF-κB	Nuclear factor-κB (including p50 and p65)
Ni-NTA	Nickel-nitrilotriacetic Acid
NMR	Nuclear magnetic resonance
NOD2	Nucleotide-binding oligomerization domain-containing protein 2
NOE	Nuclear overhauser effect
NOESY	Nuclear overhauser effect spectroscopy
Npl4	Nuclear protein localization 4 homolog
NTD	N-terminal domain
NYN	N4BP1, YacP-like Nuclease domain
NZF	Npl4 zinc finger domain
OD_{600}	Optical density at 600 nm
ODCase	Orotidine 5'-phosphate decarboxylase
ORF	Open reading frame
p53	Tumor suppressor p53
PARKIN	Parkinson juvenile disease protein 2
PCET	Proton-coupled electron transfer
PCR	Polymerase chain reaction
PDB	Protein data bank
PGN	Peptidoglycan
pH	Potential of hydrogen
PH	Pleckstrin-homology domain
pI	Isoelectric point
PIN	PilT N-terminus domain

PINK1	PTEN induced putative kinase 1
PMSF	Phenylmethanesulfonyl fluoride
PRMT	Protein arginine methyltransferases
PUB	Peptide:N-glycanase/UBA or UBX-containing protein
QH_2	Ubiquinol (dihydroxyquinone)
RABEX5	Rabaptin-5-associated exchange factor for Rab5
RAP80	Receptor-associated protein 80
RBR	RING-between-RING
RBX1	RING-box protein 1
Rel-B	Transcription factor RelB
RF	Radio frequency
RING	Really interesting new gene
RIP1	Receptor interacting protein 1
RIP2	Receptor interacting protein 2
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
RNF	RING finger protein
ROS	Reactive oxygen species
R.T.	Room temperature
RU	Resonance unit
RUB	Related to ubiquitin
SASA	Solvent accessible surface area
SCF complex	Skp, cullin, F-box containing complex
SDM	Site-directed mutagenesis
SDMA	Symmetric di-methylarginine
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SDT	Sodium dithionite
SEC	Size-exclusion chromatography
Set1/COMPASS complex	Complex proteins associated with Set1
SHARPIN	SHANK-associated RH domain interacting protein
SHE	Standard hydrogen electrode
Skp1	S-phase kinase-associated protein 1
Skp2	S-phase kinase-associated protein 2
SMG6	SMG6 nonsense mediated mRNA decay factor
Smurf1	SMAD ubiquitination regulatory factor 1
SPR	Surface plasmon resonance
SQSTM1	Sequestosome-1
STING	Stimulator of interferon genes protein
SUMO	Small ubiquitin-related modifier
TAB	TAK1-binding protein
TAE buffer	Tris-acetate-EDTA buffer
TAK1	TGF-beta-activated kinase 1
TBK1	TANK-binding kinase 1

TCA	Tricarboxylic acid cycle
TCEP-HCl	Tris (2-carboxyethyl) phosphine hydrochloride
TCR	T-cell receptor
TF	Trigger factor
TGN	*trans*-Golgi network
TLR4	Toll-like receptor 4
TNF	Tumour necrosis factor
TNFR	Tumour necrosis factor receptor
TOLIP	Toll-interacting protein
TRADD	TNF-R1-associated death domain
TRAF	TNF receptor-associated factors
Tris	Tris (hydroxymethyl) aminomethane
TXNIP	Thioredoxin-interacting protein
Ub	Ubiquitin
UBA	Ubiquitin-associated domain
UBAN	Ubiquitin binding in ABIN and NEMO
UBC	Ubiquitin-conjugating domain
Ubc13 (or UBE2N)	Ubiquitin-conjugating enzyme E2 N
UbcH5 (or UBE2D1)	Ubiquitin-conjugating enzyme E2 D1
UbcH7 (or UBE2L3)	Ubiquitin-conjugating enzyme E2 L3
UbcH10 (or UBE2C)	Ubiquitin-conjugating enzyme E2 C
UBD	Ubiquitin binding domain
Ube1	Ubiquitin-like modifier-activating enzyme 1
Ube2s	Ubiquitin-conjugating enzyme E2 S
UBL	Ubiquitin-like domain/protein
UBM	Ubiquitin-binding motif
UBP	Ubiquitin-specific-processing protease
Ub-ProT	Ub chain Protection from Trypsinization
UBR5	Ubiquitin Protein Ligase E3 Component N-Recognin 5
UCHL1	Ubiquitin C-terminal hydrolase L1
uH2A	Ubiquitinated histone H2A
uH2B	Ubiquitinated histone H2B
UIM	Ubiquitin-interacting motif
UPD	Unique Parkin domain
UPS	Ubiquitin-proteasome system
UQ	Ubiquinone
UQH_2	Reduced ubiquinone
URA3	Uridine 5'-monophosphate synthase
Urm1	Ubiquitin-related modifier 1
UV	Ultraviolet
VDAC1	Voltage-dependent anion channel 1
VPS9	Vacuolar protein sorting-associated protein 9
WT	Wild-type
XIAP	X-linked inhibitor of apoptosis protein

Y2H	Yeast-two hybrid
Y. lipolytica	*Yarrowia lipolytica*
YPD	Yeast extract–peptone–dextrose medium
ZAP	Zinc finger antiviral protein
Zap-70	70 kDa zeta-chain associated protein
ZF (or ZnF)	Zinc finger domain
ZnF216	Zinc finger protein 216
µg	Micro-gram
µL	Micro-litre
µM	Micro-molar
53BP1	TP53-binding protein 1

List of figures

CHAPTER I Introduction

CHAPTER II NMR theory

CHAPTER III Materials and Methods

CHAPTER IV Results and Disucussion

CHAPTER IV Summary and Conclusions

APPENDIX

List of tables

CHAPTER I

Introduction

Preface

Ubiquitination is among one of the most important post-translational modifications in eukaryotes. Throughout the last half century, the studies on the topic of ubiquitination has massively flourished since the discovery of ubiquitin in 1975 (Goldstein et al., 1975). The most well-known function of ubiquitination is its key involvements in protein degradation which helped Aaron Ciechanover, Avram Hershko and Irwin Rose win the Nobel Prize in Chemistry in 2004. Apart from protein degradation, ubiquitination is also widely involved in a large diversity of cellular processes, including immune signaling, protein trafficking, DNA damage responses and autophagy. Dysregulation of ubiquitination could lead to the development of multiple diseases, such as neurodegeneration, autoimmune disorder and cancer (Popovic et al., 2014). Thus, a better understanding of the functional roles in the complex ubiquitination system will greatly help and instruct the development of novel drugs and clinical treatments for our whole wellbeing.

This thesis focuses on characterizing a negative regulator N4BP1 (NEDD4-binding protein 1) in linear ubiquitination. Linear ubiquitination is mediated by the linear ubiquitin assembly complex (LUBAC) which plays crucial roles in the initiation of the innate immune response and the activation of the NF-κB signaling pathway (Gerlach et al., 2011; Ikeda et al., 2011; Tokunaga et al., 2011). The involvement of N4BP1 in the NF-kB pathway has been reported but lacks in-depth investigation. Recently, unpublished cell biological studies have demonstrated that N4BP1 is a ubiquitin-binding protein and acts as a novel regulator by inhibiting LUBAC under pro-inflammatory conditions (Kliza et al., in preparation).

To this end, this chapter will provide an overall introduction on the topics of ubiquitin, ubiquitination, LUBAC and N4BP1 respectively.

1.1 Ubiquitin and ubiquitin-related proteins/domains

1.1.1 Ubiquitin

Ubiquitin is a small protein that contains only 76 amino acids, with a molecular weight at only ~ 8.5 KDa. It is absent in eubacteria and archaea, but highly conserved in eukaryotic organisms. There is only 3-residue difference between human and yeast ubiquitin, sharing ~ 96% similarity (Figure 1.1).

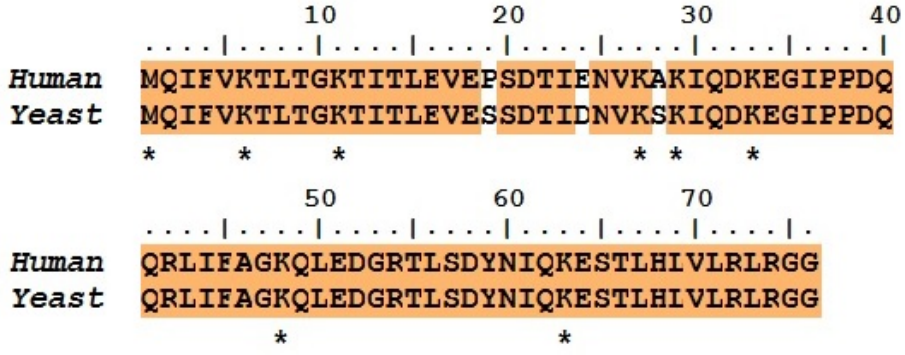


Figure 1.1 Sequence alignments of ubiquitin between human and yeast. The identical resides are highlighted in orange. Met1 and the rest of the seven Lys residues are indicated with asterisks (*).

Ubiquitin has a compact structure specified as “β-grasp fold”, consisting of five β-sheets and one α-helix, with multiple intramolecular hydrogen bonds to stabilize the structure. The carboxyl of Gly76, the last residue of the ubiquitin, can be covalently attached to the lysine residue of the substrate with the help from several enzymes to achieve ubiquitination. L8, I44 and V70 form the canonical “hydrophobic patch”, the most common interface used in ubiquitin to non-covalently interact with proteins containing ubiquitin binding domains (UBDs). Other less commonly used interfaces are also found to be critical for the interaction with UBDs, including regions centred on F4 or I36 or D58 (Figure 1.2) (Sloper-Mould et al., 2001; Haririnia et al., 2008).

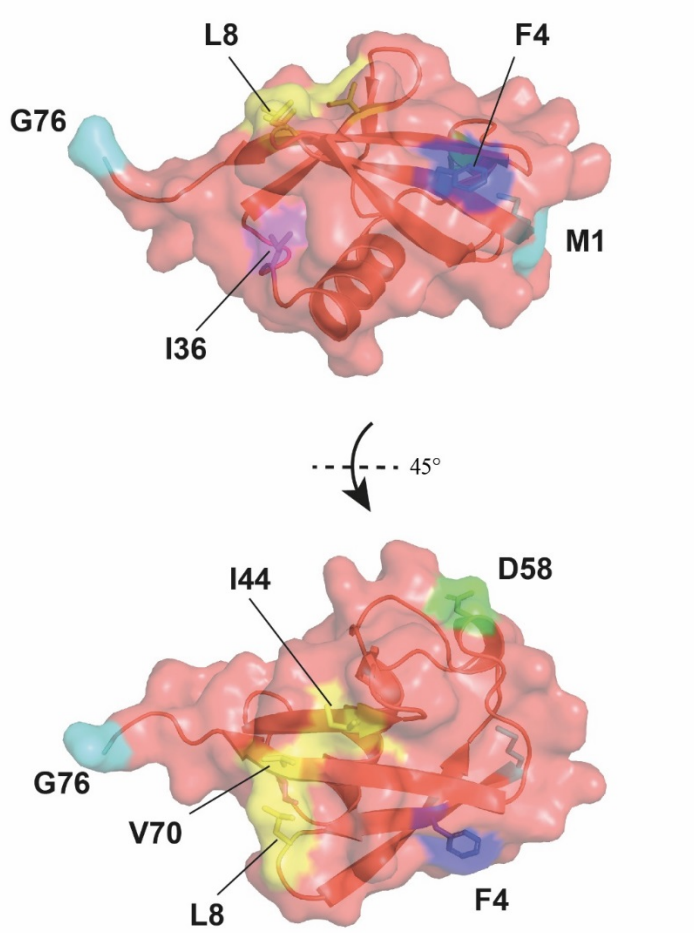


Figure 1.2 The interfaces of ubiquitin for the interaction with UBDs. Ubiquitin is shown in the surface mode. The interfaces centered on F4, I36, I44 and D58 are highlighted respectively. Figure was generated using Pymol using PDB: 1UBQ.

Throughout ubiquitination, a large variety of ubiquitin chains can be formed. There are seven lysine residues in the ubiquitin molecule (K6, K11, K27, K29, K31, K48 and K63) of which the amino group from the side chain forms an isopeptide bond with Gly76 from another ubiquitin molecule, thus elongating the ubiquitin chains with a certain inter-ubiquitin linkage. In addition, the first residue Met1 from one ubiquitin was found to be able to form a peptide bond with Gly76 from another ubiquitin molecule. Thus, this head-to-tail chain elongation produces the so called M1-linked ubiquitin chains or linear ubiquitin chains (Figure 1.3). A ubiquitin chain can be assembled in a single linkage described as “homogenous chain” or mixed with

different linkages to form the “hybrid chain” (described in detail in section 1.1.2.2 and Figure 1.5).

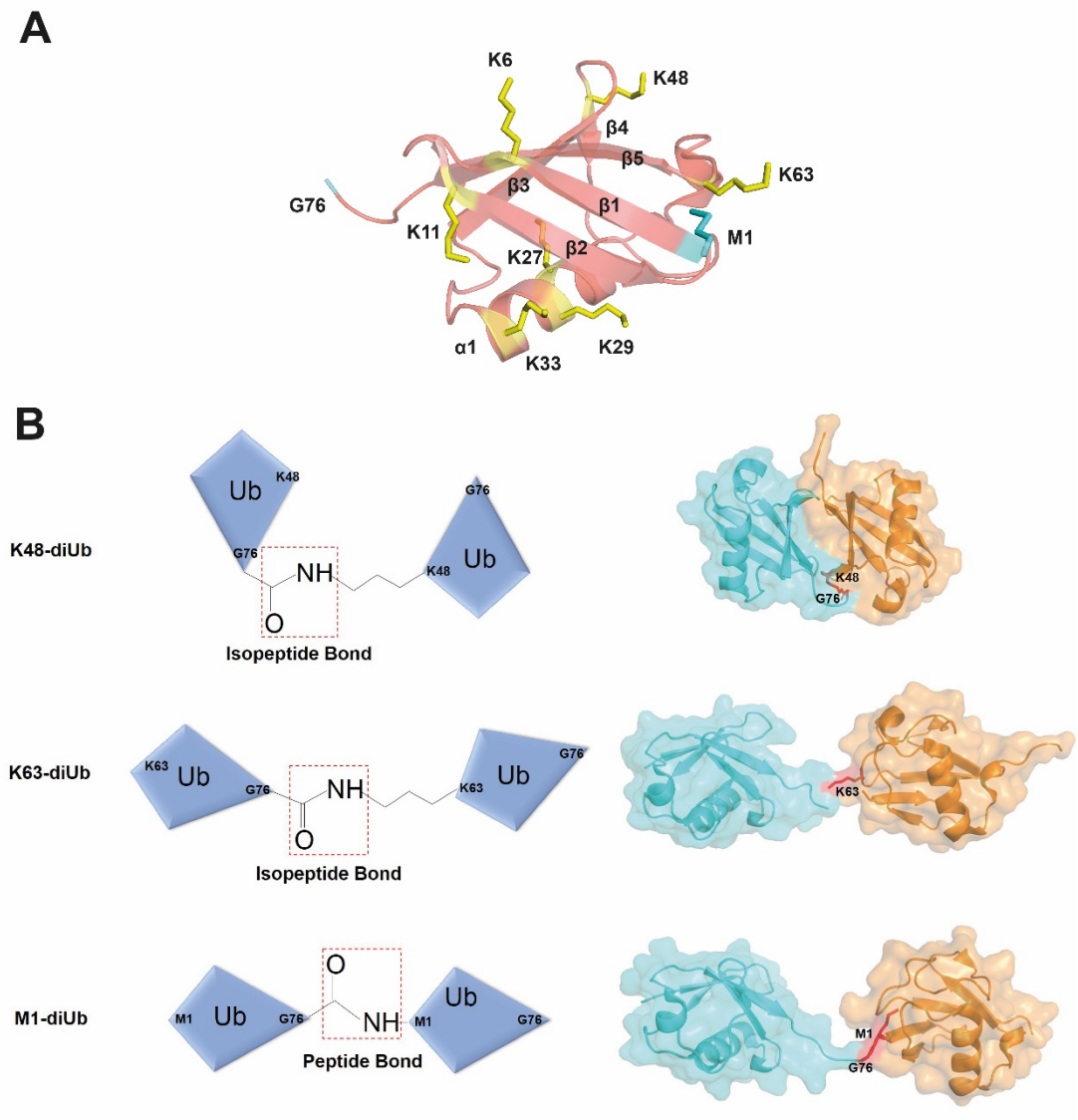


Figure 1.3 The structure of representative ubiquitin chains. A. The structure of ubiquitin where M1 and seven lysine residues involved in chain formation are highlighted. B. The structure of K48-, K63- and M1-linked di-ubiquitin chains. Ubiquitin is pictured in a kite shape to display the chain organization. The proximal ubiquitin is coloured in orange and the distal ubiquitin is in cyan. Note that G76 is absent in the crystal structure of distal ubiquitin in K63-diUb. Figure generated by Pymol using PDB: 1AAR (K48-diUb), PDB: 3H7P (K63-diUb), PDB: 2W9N (M1-diUb).

From the structural view, ubiquitin chains with different linkage types show diverse conformations. For example, the K48-linked di-ubiquitin adopts a closed conformation in which the two ubiquitin moieties interact via the I44 patch (Trempe et al., 2010). Whereas the K63- and M1-linked di-ubiquitin adopt a highly extended and flexible structure which is usually described as "open conformation" (Komander and Rape, 2012; Weeks et al., 2009) (Figure 1.3). These different linkages allow ubiquitin chains to participate and function in a variety of biological processes and pathways via modifying a wide spectrum of substrates.

1.1.2 Ubiquitin-like proteins/domains (UBLs)

1.1.2.1 Ubiquitin-like proteins

Apart from ubiquitin, there are several ubiquitin-like proteins adopting the same β-grasp topology and widely distributed in eukaryotes to regulate various biological processes. These ubiquitin-like proteins include ISG15 (Interferon Stimulated Gene 15), NEDD8 (Neural Precursor cell-expressed developmentally downregulated 8), SUMO (Small Ubiquitin-related Modifier), Atg8 (Autophagy gene 8) and Urm1 (Ubiquitin-related Modifier 1) etc.

ISG15 is the first identified UBL. ISG15 is found being strongly induced by stimulation from type I interferon and involved in innate immunity in response to viral infections. Mature ISG15 is ~ 17KDa and contains two domains which both displays the ubiquitin-like fold (Jeon et al., 2010).

There are four SUMO proteins in vertebrates but only one in yeasts and invertebrates. Though SUMO shares poor sequence similarity to ubiquitin, strikingly it still adopts the same ubiquitin β-grasp fold. Sumoylation has been found to regulate protein stability and redistribution, transcription, DNA repair etc (Yeh et al., 2000; Van der Veen and Ploegh, 2012).

Among UBLs, NEDD8 (also known as RUB in yeast) shares the highest sequence similarity with ubiquitin (58% identity). The best known NEDD8 substrate is the Cullin protein family that plays as the scaffold in the complex Cullin/RING ubiquitin ligases (CRLs). Neddylation of Cullins promotes the E3 ligase activity of CRLs. CRLs ubiquitinates a large amount of substrates, most of which are involved in the regulation of cell cycle (Van der Veen and Ploegh, 2012).

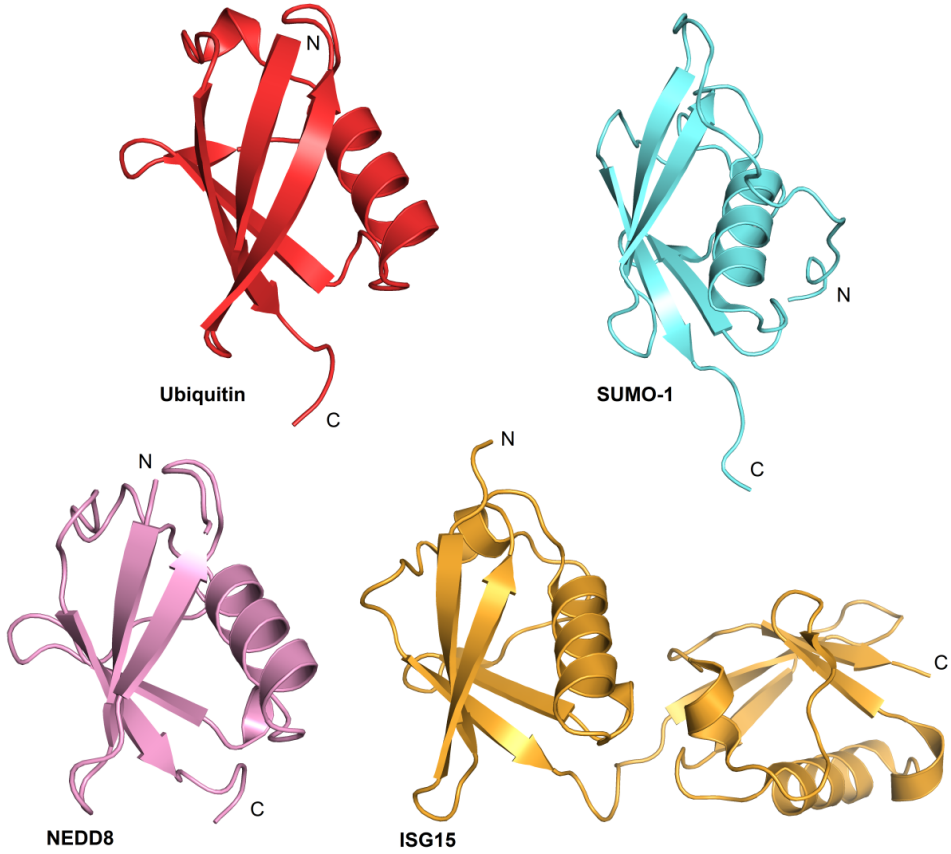


Figure 1.4 Structures of several representative ubiquitin-like proteins reveal the conserved β-grasp fold. Figure generated by Pymol using PDB: 1UBQ (Ubiquitin), 1NDD (NEDD8), 2N1V (SUMO-1), 5TLA (ISG15).

The small ubiquitin-like proteins, together with ubiquitin have greatly enlarged the variety of the post-translational modifications via targeting an enormous number of substrates. Physiologically, many proteins are tightly regulated by multiple modifications. Taking the tumor repressor p53 as an

example, p53 can be regulated by either ubiquitin or ubiquitin-like proteins. MDM2, a RING-type E3 ligase, predominantly governs the poly-ubiquitination of p53 and thus mediates its degradation via proteasome (Freedman et al, 1999). MDM2 also regulates the nuclear export of p53 by mono-ubiquitinating p53 (Li et al., 2003). NEDD8 inhibits p53 transcriptional activity via modifying MDM2 (Xirodimas et al., 2004). It is also reported that p53 activity is enhanced while modified with SUMO (Gostissa et al., 1999).

1.1.2.2 Ubiquitin-like domains

Similar to the ubiquitin-like proteins, the ubiquitin-like domains share high structure homology to ubiquitin, despite the low sequence conservation. But compared to the ubiquitin-like proteins that exist individually as small covalent effectors, UBLs are integrated into various multi-domain proteins as a certain functional unit (Walters et al., 2004; Grabbe and Dikic, 2009).

UBLs adopt a similar hydrophobic pocket to interact with binding elements such as the UBA domain, but usually less efficient compared to that of ubiquitin (Figure 1.5). This might come from the different conformation in the β1-β2 loop between UBL and ubiquitin. For example, the equivalent ubiquitin L8 located in this loop is absent in the case of $Dsk2^{UBL}$ (PDB: 2BWF), resulting in less hydrophobic contact for the interaction. Therefore, the binding affinity of $Dsk2^{UBL}$/$Dsk2^{UBA}$ is ten-fold less than Ub/$Dsk2^{UBA}$ (Lowe at al., 2005).

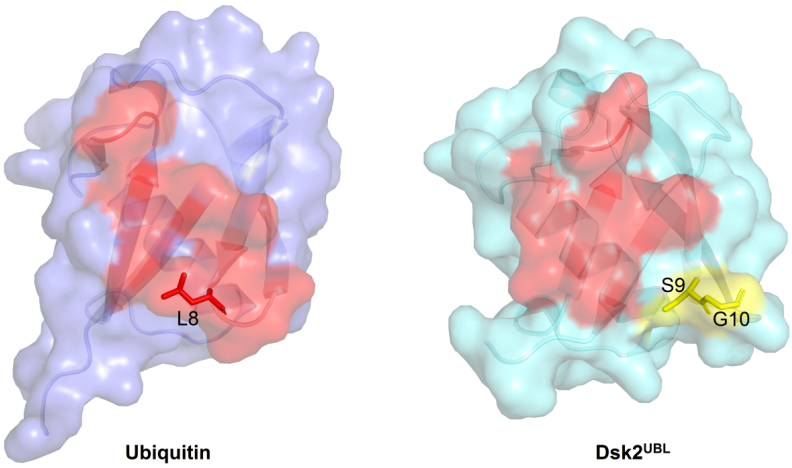


Figure 1.5 The respective UBA-binding surface on ubiquitin and Dsk2^{UBL}. The surface involved in the binding to Dsk2^{UBA} on each molecule is highlighted in red according to the comparison summarized by Lowe et al., 2005. The β1-β2 loop in Dsk2^{UBL} that is not involved in the binding is indicated by the residues S9 and G10 highlighted in yellow. Figure generated by Pymol using PDB: 1UBQ (Ubiquitin), 2BWF (Dsk2^{UBL}).

Other well-characterized UBLs include HOIL-1L^{UBL} and SHARPIN^{UBL} in the regulation of ligase activity of HOIP from LUBAC. It will be discussed in more details in section 1.3.2.

1.1.3 Ubiquitin-binding domains (UBDs)

Ubiquitin-binding domains (UBDs) are often found in proteins involved in the ubiquitin-related processes. To read the diverse "ubiquitin codes", many proteins employ UBDs to conduct ubiquitin-mediated effect. So far, more than 20 UBD families have been discovered and grouped into four major superfamilies based on their structural folds, including α-helix, Zinc finger, Ubc-like, PH domain family etc (Table 1.1).

Table 1.1 List of currently known UBDs (© 2020 Kliza and Husnjak, by permission).

Domains	Abbreviation	Name	Examples (proteins with specific UBD)
Helical	UIM	Ub-interacting motif	RPN10, VPS27, USP28, ATAXIN-3, EPS15, STAM1, STAM2, RAP80, DNAJB2, USP37, USP25, EPSINs
	MIU	Motif interacting with Ub	Rabex-5, RNF168
	UMI	UIM- and MIU-related UBD	RNF168
	DUIM	Double-sided UIM	HRS
	UBA	Ub-associated domain	PLIC1/2, HHR23A/B, p62, NBR1, Cbl-b, USP5, UBC1, HERC2, Vps13D, USP25
	CUE	Coupling of Ub to ER degradation domain	Cue2, Vps9
	GAT	GGA and TOM domain	GGA1, GGA2, GGA3, TOM1
	UBAN	Ub-binding domain in ABINs and NEMO	ABIN1, ABIN2, ABIN3, NEMO, OPTN
	VHS	VPS27, HRS and STA domain	VPS27, HRS, STAM1, STAM2, GGA1, GGA2, GGA3, TOM1
	UBM	Ub-binding motif	Polymerase iota, Rev1
	MyUb	Myosin VI UBD	Myosin VI
	AnkUBD	Ankyrin (Ank) repeat UBD	TRABID
Zinc finger (ZnF)	UBZ	Ub-binding ZnF domain	TAX1BP1, Polymerase eta, WRNIP1, FAAP20
	NZF	Npl4 ZnF domain	Npl4, Vps36, TAB2, TAB3, HOIP, HOIL-1L, SHARPIN
	ZnF A20	ZnF of A20 domain	A20, Rabex-5
	ZnF UBP (PAZ, BUZ)	ZnF of Ub-specific processing protease domain	USP5, USP20, HDAC6, BRAP2
Ub-conjugating-like	UBC	Ub-conjugating domain	UbcH5c
	UEV	Ub E2 variant domain	TSG101, Mms2
Pleckstrin-homology (PH)	GLUE	GRAM-like Ub-binding in EAP45 domain	Eap45
	PRU	Pleckstrin-like receptor for Ub	Rpn13
Others		Jab1/MPN domain	Prp8
	PFU	PLAA family UBD	Doa1, PLAA
		SH3, variant	Sla1, CIN85, amphiphysin
		WD40 repeat β-propeller	Doa1, PLAA, Fbxw8, Met30, WDR61, PAF, WDR5
		DC-UbP_N	UBTD2
		MDA-9 UBD	MDA-9

UBDs usually interact non-covalently with ubiquitin or ubiquitinated substrates in a transient manner, which allows the UBD-containing proteins to act in a rapid and timely way in response to certain stimulus. Due to large structural heterogeneity of UBDs, the binding mechanisms between UBD and ubiquitin show large multiplicities. Thus, the affinity between UBD and ubiquitin can vary widely, ranging from ~2 to 500 μM. Most of the UBDs bind to the canonical I44 patch of ubiquitin via hydrophobic interactions whereas some UBDs bind to other less common surfaces of ubiquitin (eg. the F4, L8 and I36 patches etc). Moreover, the structural diversity allows some of the UBDs to recognize specific ubiquitin chains (Figure 1.6) (Dikic et al., 2009; Rahighi and Dikic, 2012; Kliza and Husnjak, 2020). Several examples will be discussed in the next section.

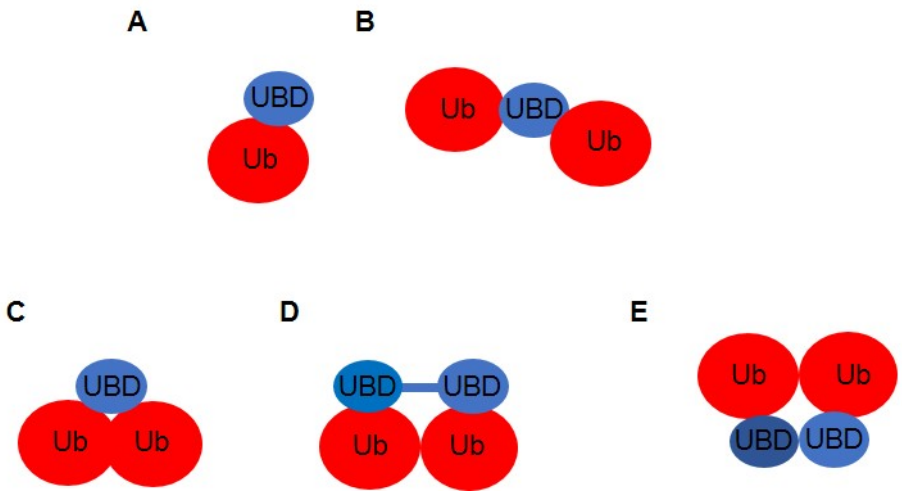


Figure 1.6 Schematic representations of different ubiquitin-binding modes by UBDs. A. One molecule of UBD binds to one molecule of mono-ubiquitin. B. One molecule of UBD binds to two molecules of mono-ubiquitin on different sites. C. A single UBD binds to ubiquitin chains. D. Two tandem UBDs from the same protein bind to ubiquitin chains. E. Two UBDs form a dimer to interact with ubiquitin chains.

1.1.3.1 The interaction between UBDs and mono-ubiquitin

The ubiquitin I44 hydrophobic patch is the most commonly used surface for the interaction with UBDs. The binding surface from UBDs can be diverse due to their multiple protein folds. Here are some examples.

The CUE (coupling of ubiquitin conjugation to ER degradation) domain belongs to the α-helix superfamily, the largest class of UBD superfamilies. It was initially named after the yeast Cue1p (Ponting, 2000). Domains from this family have been well-characterized with a compact structure comprising a three-helix bundle. The solution NMR structure of the CUE domain from the RING-type E3 ligase gp78 (also known as AMFR or RNF45) in complex with mono-ubiquitin (PDB: 2LVO) shows that the hydrophobic surface formed by α1 and α3 helices from the CUE domain is involved in the binding to ubiquitin. The conserved FP motif located in between α1 and α2 is crucial for both maintaining the structure in the appropriate fold and the hydrophobic interaction with the ubiquitin I44 patch (Figure 1.7A). Apart from it, the gp78

CUE domain binds to the proximal or distal ubiquitin from either K48- or K63-linked di-ubiquitin chains in a similar fashion as CUE/mono-ubiquitin and thus has no binding preference to K48- or K63-linkages. The ITC study confirmed that the gp78 CUE domain binds to monoUb or K48- or K63-linked di-ubiquitin chains all with comparable Kd of 12.7-14.5 μM (Liu et al., 2012).

The solution structure of the CUE2-1 domain from yeast Cue2 protein in complex with mono-ubiquitin has been solved (PDB: 1OTR), but the interaction has much lower affinity (Kd value of 155 μM) than that of gp78^{CUE}/ubiquitin. Similarly, the CUE2-1 domain mainly binds to the well-known hydrophobic I44 patch from ubiquitin but interestingly with some extension to residue K48 (Figure 1.7B). The ubiquitin K48A mutation completely abolishes the binding ability by the pull-down assay, which suggests a possible role that the CUE2-1 domain may play in inhibiting the formation of K48-linked ubiquitin chains (Kang et al., 2003).

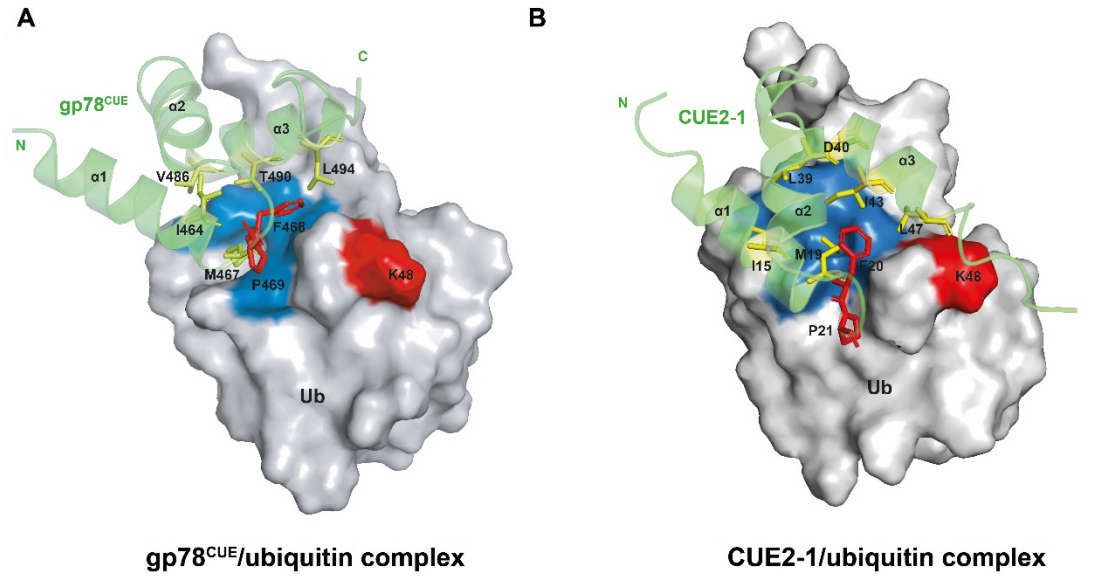


Figure 1.7 Structures of mono-ubiquitin in complex with gp78^{CUE} (A) or CUE2-1 (B). gp78^{CUE} and CUE2-1 are shown in green helices. The residues involved in the binding interface are highlighted in yellow. The FP motif is colored in red. Ubiquitin is shown in the surface mode in light gray while the I44 hydrophobic patch is in blue and K48 is in red.

Figure generated by Pymol using PDB: 2LVO (gp78^{CUE}/ubiquitin complex), PDB: 1OTR (CUE2-1/ubiquitin complex).

The NZF domain (*N*pl4 *z*inc *f*inger domain) is catalogued into the second largest class of UBDs, the Zinc finger superfamily. NZF domain is formed by four short antiparallel β–strands in which one Zn ion is coordinated by four residues (Cys or His). As exemplified by the interaction between the Npl4 NZF domain and mono-ubiquitin, the NZF domain binds to the I44 hydrophobic patch of ubiquitin via its TF/Φ motif (where Φ is any hydrophobic residue, usually Met) as the canonical ubiquitin-binding site (Figure 1.8A). The disassociation constant of the interaction is 126 μM (Alam et al., 2004; Wang et al., 2003).

Apart from the above ubiquitin I44 patch used for the binding, there are several exceptions where in some cases UBDs bind to the non-canonical areas of ubiquitin. For instance, the A20 ZnF domain from a Rab effector protein RABEX5 interacts with the D58-centred patch of ubiquitin with a Kd value of ~22 μM. A diaromatic patch consisted of Y25 and Y26 from RABEX5 governs the main interaction (Figure 1.8B) (Lee et al., 2006). Whereas the ZnF UBP domain from the deubiquitinating enzyme IsoT forms a deep hydrophobic pocket to accommodate the C-terminus of ubiquitin. In addition, F224 in the tip of the loop L2A from this domain interacts with L8 and I36 of ubiquitin. The binding has a high affinity (Kd = 2.82 μM) and is crucial for IsoT to specifically recognize the unanchored ubiquitin (Figure 1.8C) (Reyes-Turcu et al., 2006).

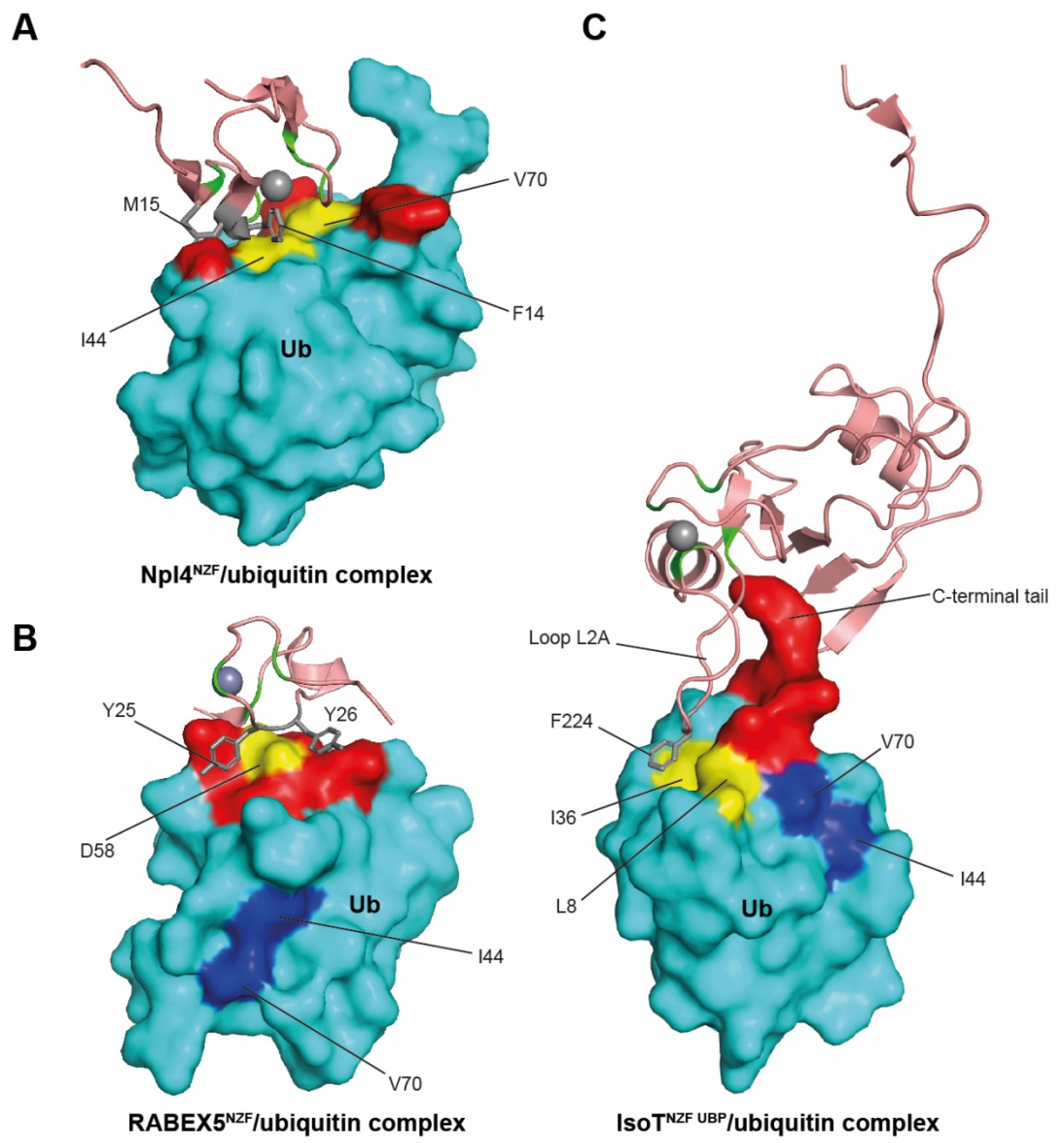


Figure 1.8 Structures of the Npl4 NZF domain (A), the RABEX5 NZF domain (B) and IsoT NZF UBP domain (C) in complex with mono-ubiquitin respectively. Ubiquitin is shown in the surface mode in cyan while the UBDs are shown in the cartoon mode. The binding surface of ubiquitin is coloured in red and yellow. The key residues from UBDs involved in the binding are shown as sticks in grey. The Zn ion is shown as a grey sphere and coordinated by four Cys/His residues highlighted in green. Figures were generated by Pymol using PDB: 1Q5W (Npl4^{NZF}/ubiquitin complex), PDB: 2FIF (RABEX5^{NZF}/ubiquitin complex) and PDB: 2G45 (IsoT $^{NZF\ UBP}$/ubiquitin complex).

The above examples describe the conditions where UBD simply binds to mono-ubiquitin in a 1:1 ratio. However, the HRS UIM domain displays a double-sided ubiquitin-binding mode where one molecule of the UIM domain could bind to both I44 patches of two mono-ubiquitin molecules. The ubiquitin-binding occurs simultaneously in the opposite sides of this domain (Figure 1.9). SPR study revealed that the disassociation constant of the interaction is 190 μM (Hirano et al., 2006).

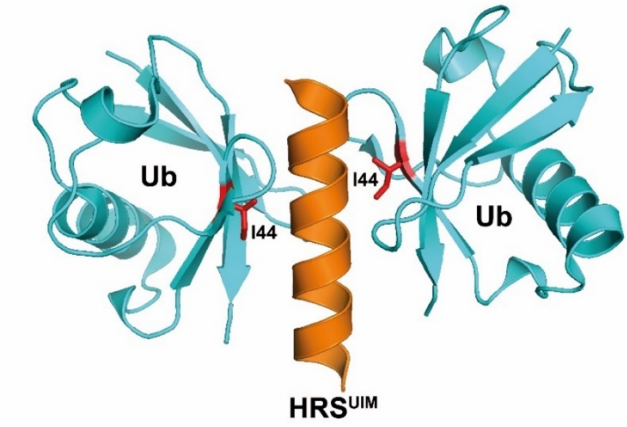


Figure 1.9 Structures of the HRS NZF domain in complex with mono-ubiquitin. Ubiquitins are coloured in cyan while the HRS UIM domain is in orange. I44 residues are shown as red sticks. Figure was generated by Pymol using PDB: 2D3G.

On the ubiquitin side, its multiple contact surfaces could also bind to more than one UBD at a time. For example, the ubiquitin receptor ZnF216 and SQSTM1 are both involved in shuttling ubiquitinated proteins to the 26S proteasome. A ternary complex model consisted of the two UBDs and mono-ubiquitin is proposed (Garner et al., 2011). NMR titrations demonstrated that the A20 ZnF domain of ZnF216 binds to the D58-centred patch of ubiquitin while the UBA domain of SQSTM1 binds to the I44 patch of ubiquitin. Thus, ubiquitin is competent to act as a scaffold to recruit multiple UBDs, which may facilitate the complex assembly.

1.1.3.2 The binding specificities of UBDs to ubiquitin chains

Apart from the interaction with mono-ubiquitin, some of the UBDs display preference to certain types of ubiquitin chains. Due to the multiplicity of both UBDs and ubiquitin chains, the specificity determinants between the two binding partners are largely diverged. For some of the UBDs, a single domain is capable of recognizing a certain ubiquitin chain type while in other cases multiple UBDs cooperatively work together to display chain specificity. In addition to the contact with each ubiquitin moiety, some of the UBDs could also recognize the linker region from the ubiquitin chain.

1) Ubiquitin-binding by a single UBD

There are observations where a single UBD has multiple ubiquitin-binding sites to interact with a certain type of ubiquitin chain.

TAB2, for example, plays as an adaptor protein binding to K63-linked ubiquitin chains to activate the TAK1 kinase complex, thus facilitating the activation of the downstream NFκB signaling (Kanayama et al., 2004). The NZF domain from TAB2 binds to both I44 hydrophobic patches from the two ubiquitin moieties of K63-linked di-ubiquitin. The TF/Φ motif from this domain is employed as the canonical ubiquitin-binding site to interact with the distal ubiquitin while three conserved residues from the side of the domain are responsible for the contact with the proximal ubiquitin (Figure 1.10A). The TAB2 NZF domain shows distinct binding preference to the K63 linkage (Kd value of 8 μM for K63-diUb) rather than the linear linkage (Kd value of 148 μM for M1-diUb). This specificity lies in that the linear chains cannot bend in a similar way to that of the K63 linkage, therefore TAB2 is characterized to bind only one ubiquitin moiety in the linear di-ubiquitin (Kulathu et al., 2009).

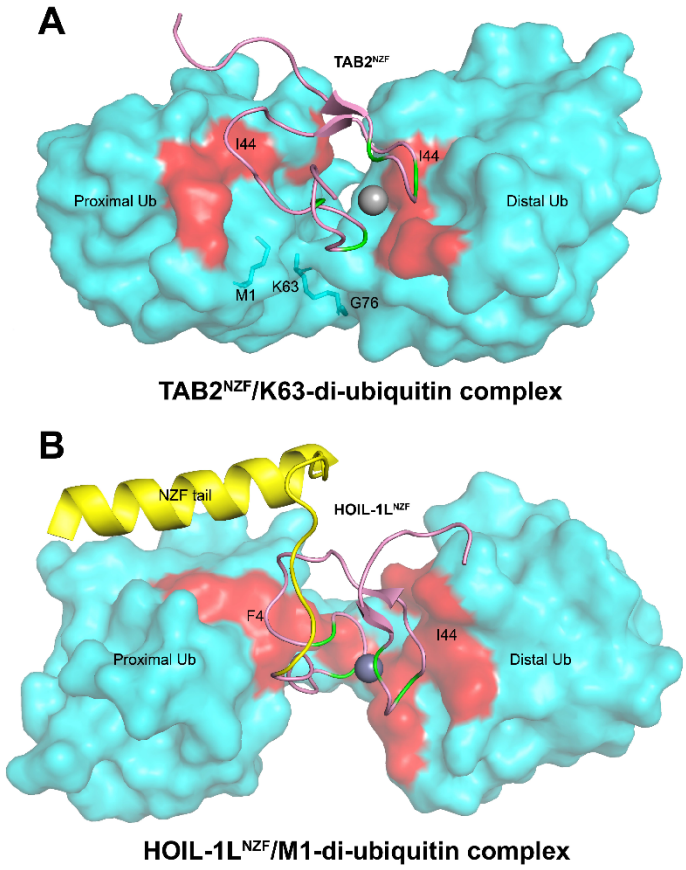


Figure 1.10 The structure of the TAB NZF domain in complex with K63-linked di-ubiquitin (A) and the HOIL-1L NZF domain in complex with M1-linked di-ubiquitin (B). Di-ubiquitin chains are shown in cyan in the surface mode where the contact surfaces are highlighted in red. The NZF domains are shown in pink, where the four Cys residues coordinating the Zn ion are indicated in green. The Zn ion is displayed as a grey sphere. The NZF tail of HOIL-1L^{NZF} is coloured in yellow. Figures were generated by Pymol using PDB: 2WWZ (TAB2^{NZF}/K63-di-ubiquitin complex), PDB: 3B08 (HOIL-1L^{NZF}/M1-di-ubiquitin complex).

The NZF domain from HOIL-1L, consisting both the "NZF core" and the unique "NZF tail", binds to linear di-ubiquitin chains with a disassociation constant of 17.2 μM, which is ~ 20-fold higher than the K48- or K63-linked di-ubiquitin chains. In the HOIL-1L^{NZF}/M1-di-ubiquitin complex (PDB: 3B08), the NZF core, which is almost structurally identical to the other canonical NZFs, interacts with the two ubiquitin moieties in a two-sided

interaction. Another element “NZF tail” offers additional support to enhance the specificity. Apart from the canonical hydrophobic binding between the NZF TF/Φ motif and the distal ubiquitin I44 patch, another part of the NZF core, as well as the NZF tail, contribute several residues which are only unique to HOIL-1L to recognize the F4-centred surface of the proximal ubiquitin. This binding mode is different from that of the $TAB2^{NZF}$/K63-di-ubiquitin complex, which explains its binding preference to the linear linkages (Figure 1.10B) (Sato et al., 2011).

2) Recognizing the linker region of ubiquitin chains

In addition to the direct contact with both ubiquitin moieties as shown above, some of the UBDs could also directly interact with the linker region of a certain type of ubiquitin chain.

As some studies direct UBA-containing proteins in the proteasomal degradation machinery (Hofmann and Bucher, 1996; Chen and Madura, 2002), there are cases where some UBA domains show binding preference to K48 linkages. The UBA2 domain from HHR23A (also known as RAD23A), for instance, forms a distinctive sandwich-like complex (PDB: 1ZO6) with K48-linked di-ubiquitin chain (Varadan et al., 2005). Helix 2 from $HHR23A^{UBA2}$ interacts with the I44 patch from the proximal ubiquitin, while helix 3 mainly binds to an extended hydrophobic pocket including the C-terminus of the distal ubiquitin and the linker area between the two ubiquitin moieties. Additionally, loop 1 of UBA2 binds to residue G47 and K48 from the distal ubiquitin to enhance the binding (Figure 1.11).

Other examples include the DUB domain from the deubiquitinating enzyme AMSH-LP, where the catalytic groove from this domain recognizes the isopeptide bond in the linker region in K63-linked di-ubiquitin. This binding mode thus explains the catalytic specificity of AMSH family to ubiquitin chains with K63 linkage (Sato et al., 2008).

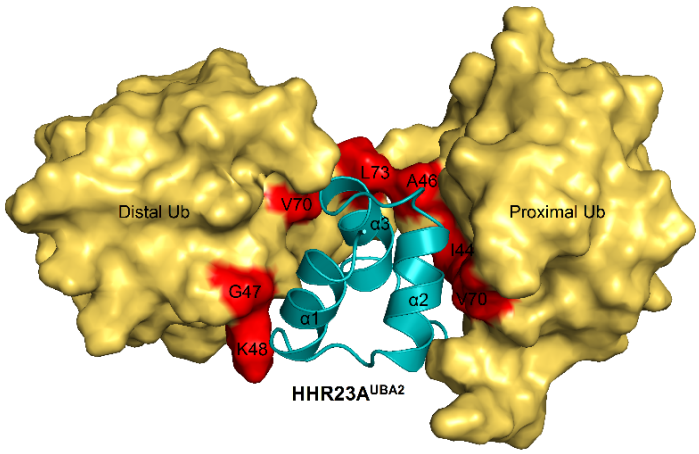


Figure 1.11 The structure of the HHR23A UBA2 domain in complex with K48-linked di-ubiquitin. HHR23A^{UBA2} is shown in light blue helices. K48-linked di-ubiquitin is shown in the surface mode in yellow. The residues involved in the binding interface are highlighted in red. Figure was generated by Pymol using PDB: 1ZO6.

3) Ubiquitin-binding by multiple UBDs

It is interesting to find that some of the ubiquitin-binding proteins contain more than one UBDs. Structural studies have demonstrated that multiple UBDs could work together in a specific spatial arrangement to recognize certain ubiquitin chain types.

The crystal structure of the tandem UIM domains from RAP80 in complex with K63-linked di-ubiquitin (PDB: 3A1Q) reveals the binding mechanism that defines the chain specificity of the protein. RAP80 UIM1 and UIM2 domains are tandemly linked via the inter-UIM region to form a continuous α-helix, where UIM1 interacts with the proximal ubiquitin while UIM2 interacts with the distal ubiquitin. Both UIM domains recognize the I44 patch of ubiquitin, but neither of them directly contacts with the linker region of K63-diUb. From the complex structure, the inter-UIM region does not recognize the linker region either. However, the length of the inter-UIM region is critical for the binding specificity in a precise way where it arranges the two UIM domains in the most appropriate positions to bind to both

ubiquitin moieties simultaneously. Experimental evidence has shown that either increasing or reducing the residues in the inter-UIM region will fully abolish the interaction. It is thus described as a “molecular ruler” as it could “measure” the distance between the two ubiquitin moieties (Figure 1.12) (Sato et al., 2009).

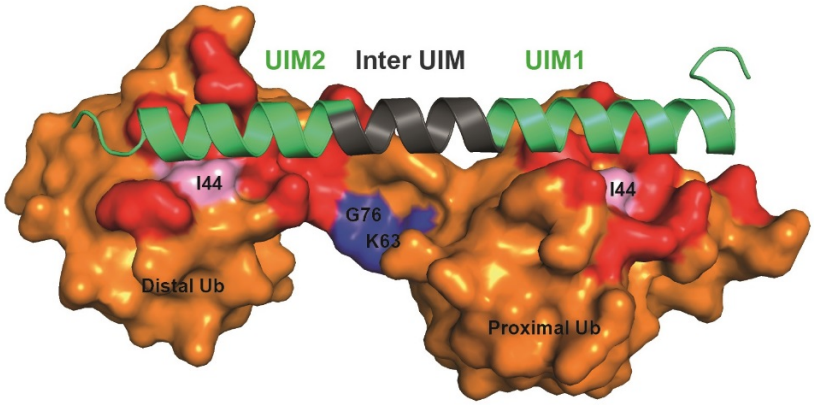


Figure 1.12 The structure of the RAP80 UIM domain in complex with K63-linked di-ubiquitin. RAP80^{UIM} is shown in light blue and grey helices. K63-linked di-ubiquitin is shown in the surface mode in orange. The residues involved in the binding interface are highlighted in red and pink. Figure was generated by Pymol using PDB: 3A1Q.

The NEMO UBAN domain has been characterized to selectively interact with M1-linked ubiquitin chains with high affinity (Kd value of 1.6 μM for M1-diUb). The structural study has shown that the UBAN domain displays a coiled-coil arrangement and forms a parallel homodimer to interact with two M1-linked di-ubiquitin molecules (PDB: 2ZVO) (Figure 1.13) (Rahighi et al., 2009). The two ubiquitin moieties in the M1-linked di-ubiquitin have distinct interaction with the UBAN dimer. The distal ubiquitin contacts with both UBAN protomers via the canonical I44 patch and the C-terminal tail (from R72 to R74). The C-terminal tail is part of the linker region, demonstrating the ability from the UBAN domain to recognize the M1 linkage. In contrast to the major hydrophobic interaction between the distal ubiquitin and UBAN, residues that are located next to the I44 patch in the proximal ubiquitin are

employed for mostly polar interactions with only one of the UBAN protomers. The I44 patch from the proximal ubiquitin is not involved in the interaction.

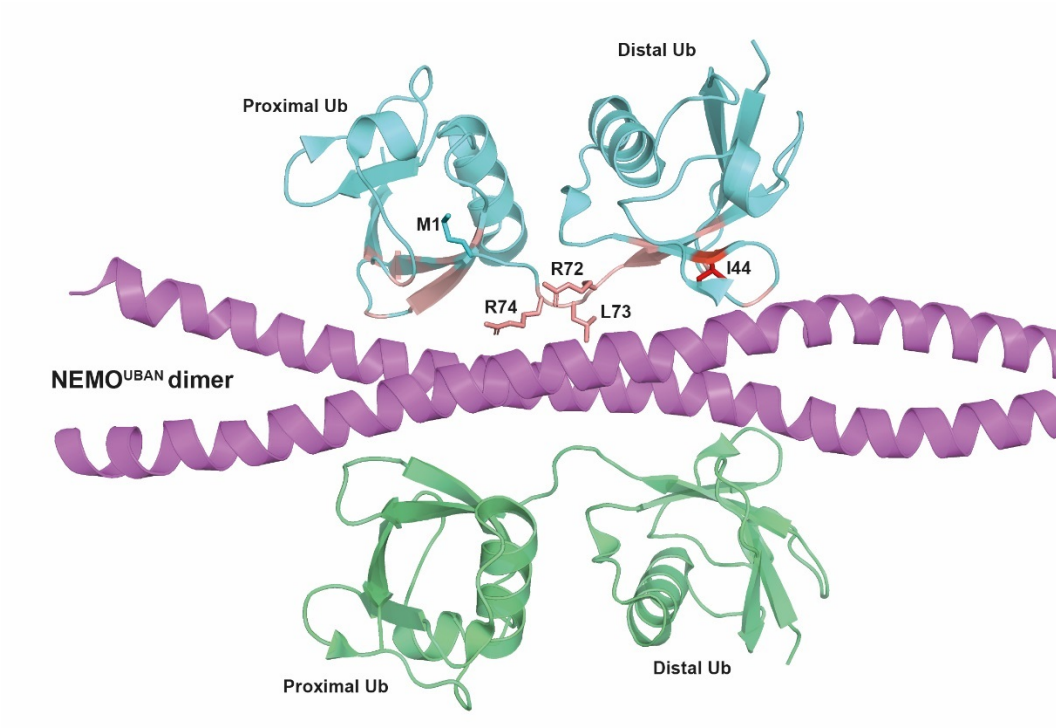


Figure 1.13 The structure of the NEMO UBAN domain in complex with M1-linked di-ubiquitin. NEMO^{UBAN} is coloured in purple and shown in the dimer form. Two M1-linked di-ubiquitin molecules are coloured in cyan and green respectively. The key residues involved in the interaction are coloured in pink and shown in one of the di-ubiquitin molecules where I44, R72-R74 from the distal ubiquitin are highlighted as sticks. Figure was generated by Pymol using PDB: 2ZVO.

1.2 Ubiquitination

1.2.1 Ubiquitination cascade

To achieve ubiquitination, three different enzymes team up in a cascade to generate the correct and precise modification of the substrate with ubiquitin (or chains) (Figure 1.14). These enzymes include the E1 ubiquitin-activating enzyme, the E2 ubiquitin-conjugating enzyme and the E3 ubiquitin ligase.

At the first step, E1 hydrolyses ATP to form a ubiquitin-adenylate intermediate (Ub-AMP) at the C-terminus of ubiquitin. Ub-AMP is then attacked by the side chain from E1 catalytic Cys, forming a high-energy thioester bond ("~" is used hereafter) between the two molecules, coupled with the adenylation of the second ubiquitin (Schulman and Harper, 2009). The activated ubiquitin from E1~Ub is subsequently delivered onto the E2 active site Cys by a transthiolation process, forming the E2~Ub conjugate (Ye and Rape, 2009). E3 ubiquitin ligases play crucial roles in substrate recognition and ligation with ubiquitin to achieve the final step of ubiquitination. Depending on their catalytic mechanism, some E3 ubiquitin ligases will accept the activated ubiquitin from E2 to form an E3-ubiquitin thioester intermediate (E3~Ub) and subsequently transfer ubiquitin to the substrate, such as the HECT- and RBR-type ligases. In contrast, RING-type E3 ligases work as a bridge between ubiquitin charged E2 and the substrate, where the ubiquitin transfer is performed directly from E2 to the substrate. (Zheng and Shabek, 2017).

In humans there are only two E1 ubiquitin-activating enzymes, 30 E2 ubiquitin-conjugating enzymes and more than 600 E3 ligases in human (Brazee et al., 2016). Therefore, different E2/E3 combination can catalyze the formation of various ubiquitin linkages to take part in a wide range of biological pathways.

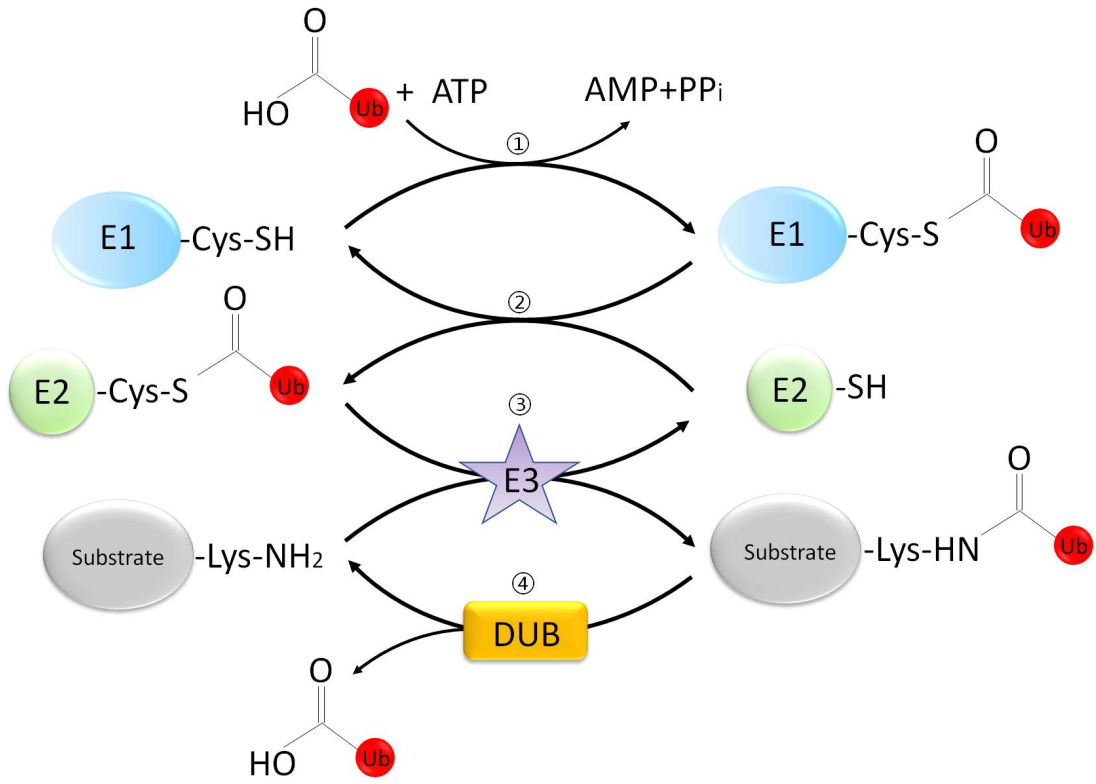


Figure 1.14 Schematic of the ubiquitination cascade and de-ubiquitination. Ubiquitination consists of 3 steps: ①: ubiquitin activation by E1; ②: E2 conjugation; ③: E3 ligation. Step ④ indicates the de-ubiquitination process catalyzed by DUB.

Ubiquitination can be reversed by a group of enzymes called deubiquitinating enzymes (DUBs) (Figure 1.14). DUBs cleave ubiquitin chains into oligomers or monomers or process the ubiquitin precursor into the mature ubiquitin. Most important of all, DUBs catalyze the removal of the ubiquitin or chains from the substrate, thus reversing its fate. Some of the DUBs can also catalyze the cleavage of ubiquitin-like proteins attached to the substrates (Ramakrishna et al., 2011; Reyes-Turcu et al., 2009).

Taking together, the dynamic ubiquitin modification is under strict and precise control by both ubiquitination and de-ubiquitination, to guarantee the intricate activities in living organisms.

1.2.2 Cellular function of ubiquitination

The versatility of ubiquitination allows its involvements in almost every aspect of the biological processes. In addition to the simple mono-ubiquitination, there are many different means of poly-ubiquitination (Figure 1.15). A single substrate can be multi-mono-ubiquitinated at several sites or modified by chains with certain linkages. Less frequently, substrates can be ubiquitinated by a hybrid of two or more linkages on one chain or even branched chain. Most of them require the specific recognition by UBDs, therefore, to greatly diversify their functions. Accumulating evidence has demonstrated that a combination and collaboration of multiple types of ubiquitin chains is needed in the precise regulation of some intricate pathways, such as the TNF-induced NFκB pathway (Rittinger and Ikeda, 2017; Spit et al., 2019). Due to relatively higher enrichment in the cells, K48 and K63 linkages are widely studied throughout decades. However, the functional roles of K6, K11, K27, K29 and K33 linkages are still lack of systematic investigation.

Apart from the linkage, the ubiquitin chain length is also an important determinant for the selective recognition by UBDs, but still poorly understood due to the lack of powerful tools. It is previously known that a minimum of K48-linked tetra-ubiquitin chain is required for the recognition by the proteasome (Pickart, 2000). Until recently, a novel biochemical method, Ub-ProT (*U*b chain *pro*tection from *t*rypsinization), was developed to determine the ubiquitin chain length from substrates. Using this method, it was found that EGFR was mainly modified with K63-linked chains of 4-6 ubiquitin moieties upon activation by EGF (Tsuchiya et al., 2018). In another study, the Ub chains with defined length and linkage were used as baits to identify ubiquitin-binding proteins in a global scale. For example, it has been shown that the K29- and K33-linked ubiquitin chains with shorter length (di-ubiquitin or tetra-ubiquitin) preferentially bind to protein-modifying

enzymes, while longer chains with at least 6 ubiquitin moieties have the tendency to bind metabolite interconversion enzymes (Lutz et al., 2020). With the emerging techniques, it will surely help us exploring the function of ubiquitin length in a lot more detail.

In the below section, A brief functional introduction in the context of only linkages will be summarized.

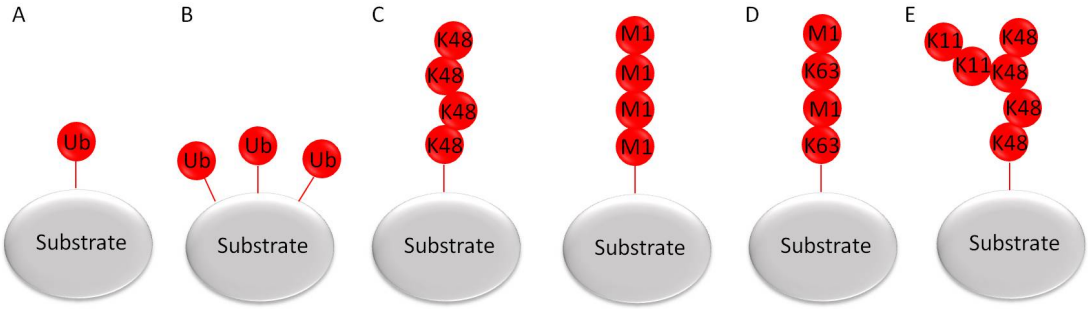


Figure 1.15 Examples of different means of ubiquitination. A. Mono-ubiquitination. B. Multi-mono-ubiquitination. C. Modification by homogenous ubiquitin chain. D. Modification by hybrid ubiquitin chain. E. Modification by branched ubiquitin chain.

Mono-ubiquitination

The studies on mono-ubiquitination are mostly centred on H2A and H2B, revealing its close association with gene expression, transcription, DNA replication and DNA repair etc. hPRC1L is an E3 ligase complex that specifically targets mono-ubiquitination on H2A and mediates Polycomb target gene silencing (Wang et al., 2004). The reduced level of mono-ubiquitinated H2A (uH2A) decreases the stability of the nucleosomes, implicating its possible roles in regulating transcriptional repression (Zhu et al., 2007; Weake and Workman, 2008).

In contrast to the negative regulatory roles of uH2A, mono-ubiquitinated H2B (uH2B) is suggested to activate gene expression in a crosstalk with histone methylation, where the mono-ubiquitination of H2B is prerequisite for the methylation of histone H3K4 and H3K79 (Sun et al., 2002; Weake and Workman, 2008). For example, the association between histone

methyltransferase complex Set1/COMPASS and transcribed chromatin is stabilized by uH2B, therefore the activity of Set1/COMPASS is facilitated to further activate gene expression (Racine et al., 2012). uH2B also promotes DNA repair via increasing the methylation of H3K4 on the DNA broken site in response to DNA damage (Faucher et al., 2010).

K48-linked ubiquitination

The K48-linked poly-ubiquitination is the most abundant ubiquitin modification. The study is centred on the ubiquitin-proteasome system (UPS). In this system, K48-linked ubiquitin chains act as "degradation tags" to covalently attach to the target protein through the ubiquitination process and mediates protein degradation via the 26S proteasome (Thrower et al., 2000; Komander and Rape, 2012). Once targeted with K48-linked chains, the substrate is recognized by the 19S regulatory particle of the proteasome and then digested into short peptides in the 20S subunit. The ubiquitin is then free from the substrate and can be reused again (Ciechanover and Schwartz, 1998).

The proteosomal degradation via K48 linkage not only eliminates undesired proteins, but also participates in several signaling pathways. For example, in the canonical NFκB pathway, the K48-mediated IκBα degradation is required in order to release the NFκB dimer p50/p65 into the nucleus, therefore initiating effector gene transcription. Similarly, in the non-canonical NFκB pathway, the p52 precursor p100 is targeted for partial proteasomal degradation and then processed into mature p52. p52 then forms a complex with Rel-B to activate gene expression in the nucleus (Liu and Chen, 2011; Pamukcu et al., 2011).

Interestingly, K48 linkage is also suggested to carry proteolysis-independent functions. As exemplified by the transcription factor Met4, CDC34/SCF mediates the ubiquitination of Met4 with K48-linked

chains. However, as a result, the transcriptional activity of Met4 is inhibited but not degraded (Kaiser et al, 2000). Later detailed mechanism reveals that the UBD domain of Met4 limits the length of K48-linked chains formed on Met4, preventing its recognition by the proteasome (Flick et al., 2006).

K63-linked ubiquitination

In contrast to the close relation between K48 linkage and proteosomal degradation, K63-linked poly-ubiquitination acts as a regulatory signal and is mostly involved in non-proteolytic processes, such as DNA damage response, autophagy, innate immunity and protein trafficking.

When recruited to the DNA damage sites, the RING-type E3 ligase RNF8, together with an E2 enzyme Ubc13, catalyzes K63-linked ubiquitin chains onto H2A and H2AX. The K63-linked chains mediate the recruitment of multiple effectors for the DNA repair, implying key roles in DNA damage response (Bennett and Harper, 2008; Chen and Sun, 2009).

It has been reported that the K63-linked ubiquitination of Beclin-1 mediated by TRAF6 regulates TLR4-induced autophagy during inflammatory responses (Shi and Kehrl, 2010).

Notably, K63 linkage participates in various processes via kinase activation. RNF128 catalyzes the K63-linked ubiquitination of kinase TBK1, thus inducing the activation of TBK1 and the subsequent IFN-β production in antiviral immune response (Song et al., 2016). TRAF6 also targets the serine/threonine protein kinase Akt with K63-linked ubiquitination, required for Akt phosphorylation and membrane recruitment (Yang et al., 2009).

Surprisingly, K63-linked ubiquitination was recently found to act as a substrate-specific marker and is indirectly involved in proteasomal degradation. K63-polyubiquitinated TXNIP catalyzed by ITCH triggers the recruitment of UBR5 while UBR5 subsequently forms K48 branches on the K63 linkage, resulting in the degradation of TXNIP (Ohtake et al., 2018).

K6-linked ubiquitination

K6-linked ubiquitin chain has the least abundance among all the linkages. Currently it is reported to take part in DNA repair and mitochondria homeostasis. Its involvements in DNA repair are mostly related to BRCA1. The heterodimer BRCA1/ BARD1 acts as an E3 ligase and auto-ubiquitinates BRCA1 with K6-linked poly-ubiquitin chains (Wu-Baer et al., 2003). It was found that BRCA1 and BRCA1-dependent K6-linked chains are recruited to the DNA damage sites, implicating their participation in DNA repair (Morris and Solomon, 2004). Another study reveals that the mutated BRCA1 lacking the E3 activity for auto-ubiquitination is relatively unstable, suggesting a role of K6 linkage in substrate stability (Reid et al., 2008).

An E3 ligase Parkin assembles K6-, K11- and K63-linked ubiquitin chains when stimulated by mitochondrial damage (Cunningham et al., 2015). Mitochondrial autophagy (mitophagy) was found to undergo serious delay when wild-type ubiquitin is replaced with Ub-K6R or Ub-K63R in the cells using a ubiquitin replacement strategy (Ordureau et al., 2015).

Recently, a HECT-type E3 ligase HUWE1 was demonstrated to modify Mfn2 with K6 linkages, but the specific function of this modification on the substrate is yet unidentified (Michel et al., 2017).

K11-linked ubiquitination

K11-linked ubiquitination is also involved in proteasomal degradation. The E3 ligase APC/C, together with two E2 enzymes UbcH10 and Ube2s, accounts for the K11-linked poly-ubiquitination on Cyclin B1, thus mediating its degradation by the proteosome to facilitate the mitosis progress. It suggests that K11-linked ubiquitination plays as a degradative signal in regulating cell cycle (Kirkpatrick et al, 2006; Jin et al, 2008; Garnett et al, 2009; Williamson et al, 2009).

Interestingly, Ube2s is also found to catalyze the K11-linked ubiquitin chain formation on β-Catenin, but β-Catenin is stabilized instead (Li et al., 2018). An E2 family, UbcH5, has been demonstrated to increase the K11-linked poly-ubiquitination on RIP1 in the TNFR1 signaling (Dynek et al., 2010).

K27-linked ubiquitination

The studies on K27 linkage reveal its possible participation in mitophagy, innate immunity and DNA damage. Parkin-mediated K27-linked ubiquitination targets VDAC1 and VDAC1 is essential for PINK1/Parkin-directed mitophagy (Geusler et al., 2010). Double-strand breaks (DSBs) induce major K27-linked ubiquitination on chromatin. RNF168 is the specific histone E3 ligase in charge of this catalysis. RNF168-dependent K27-linked ubiquitination on Histone H2A/H2A.X is required for activating DNA damage response. It has been observed that BRCA1 and 53BP1 cannot be recruited to the DNA damage sites when K27 linkage is absent (Gatti et al., 2015).

The knowledge on K27 linkage involved in innate immunity has been enlarged by recent research. In reaction to DNA stimulation from virus such as HSV-1, AMFR-INSIG1 complex acts as an E3 ligase to modify STING with K27-linked chains. It facilitates the recruitment of TBK1 by STING to form the STING-TBK1 complex, which further activates the transcriptional factor IRF-3. IRF-3 then induces the expression of type I interferons and pro-inflammatory cytokines (Wang et al., 2014). Upon HSV-1 infection, RNF185-mediated K27-linked ubiquitination on cGAS facilitates the enzymatic activity of cGAS. The formation of 2'3'-cGAMP catalyzed by cGAS acts as a second messenger to active the downstream STING signaling (Wang et al., 2017).

K29-linked ubiquitination

The function of K29 linkage in lysosomal degradation has been demonstrated. In the Notch pathway, AIP4, the human orthologue of mouse HECT-type E3 ligase ITCH, directly interacts with DTX and ubiquitinates DTX with K29-linked ubiquitin chains for lysosomal degradation (Chastagner et al., 2006). In another study, NEMO and p65 are targeted with K29-linked ubiquitin chains by TRAF7 for lysosomal degradation, suggesting that K29 linkage also participates in the NFκB signaling pathway (Zotti et al., 2011).

There are also examples that K29 linkage is related to non-proteolytic functions. Apart from catalyzing the formation of K48 linkage, Smurf1 is able to modify Axin with K29-linked chains, egatively regulating Wnt/β-Catenin signaling. As a result, Axin lost the ability to interact with the Wnt coreceptors LRP5/6 (Fei et al., 2013).

K33-linked ubiquitination

K33 linkage is suggested to play as an inhibitory role, especially in T-cell activation. It is reported that the ubiquitination of AMPK-related kinases mediated by K29 and K33 linkages inhibits their phosphorylation and activation by LKB1 (Al-Hakim et al., 2008). Two E3 ligases Cbl-b and ITCH work cooperatively to target K33-linked ubiquitination on T-cell receptor TCR-ζ. Then phosphorylation of TCR-ζ and its association with the signaling kinase Zap-70 are reduced, which further inhibits the T cell activation (Huang et al., 2010). Zap-70 was also found to be modified with K33 linkage by Nrdp1, resulting in the inactivation of Zap-70, thus terminating early TCR signaling in $CD8^+$ T cells (Yang et al., 2015).

K33 linkage is also involved in protein trafficking. CUL3-KLHL20 is an E3 ligase that catalyzes the attachment of K33-linked ubiquitin chains on Crn7, which promotes Crn7 targeting to the *trans*-Golgi network (TGN) (Yuan et al., 2014).

The functional studies on linear ubiquitination will be discussed in the next section.

1.2.3 Linear ubiquitination

To date, the linear ubiquitin assembly complex (LUBAC) is the only known E3 ligase that mediates the formation of linear ubiquitin chains. Initially, it was found that LUBAC was composed of two components HOIP (HOIL1-interacting protein, also known as RNF31) and HOIL-1L (Haem-oxidized IRP2 ubiquitin ligase-1, also known as RBCK1) (Kirisako et al., 2006). Then SHARPIN (Shank-associated RH domain-interacting protein) was discovered as part of the complex to regulate the NF-κB pathway and apoptosis (Ikeda et al., 2011). LUBAC soon drew wide attention in the ubiquitin research field.

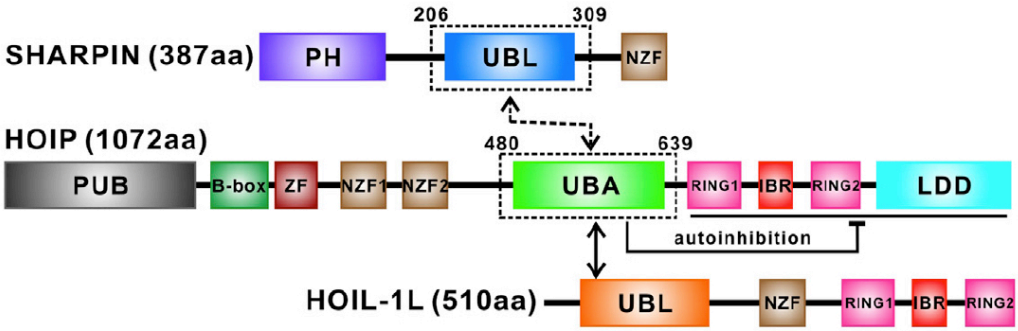


Figure 1.16 Schematic of the architectures of the three subunits of LUBAC as well as the interactions between the UBL and UBA domain (© 2017 Liu et al., by permission).

LUBAC is characterized as a 600-KDa protein complex although the exact stoichiometry of the three components is not yet defined. HOIP and HOIL-1L are both identified as RBR-type E3 ligases, of which HOIP is the main catalytic subunit in the complex (Figure 1.16). HOIL-1L and SHARPIN function as important regulators for HOIP ligase activity. Although HOIL-1L is also capable of synthesizing linear ubiquitin chain, the *in vitro* ligase activity is much weaker than HOIP (Stieglitz et al., 2012).

The mechanism of linear ubiquitin chain elongation (in the absence of physiological substrates) by HOIP has been studied by structural approaches. It will be discussed in section 1.3.2.2. In terms of linear ubiquitination of a real substrate, such as NEMO, the catalysis requires key regulators from the LUBAC complex. It has been shown that HOIL-1L is not only required for releasing the auto-inhibition of HOIP (please see section 1.3.2.1), but also facilitating HOIP to initiate mono-ubiquitination on NEMO. Rather than the en block transfer of linear ubiquitin chains to the substrate, this priming mono-ubiquitination is followed by the processive linear chain extension by HOIP (Smit et al., 2013).

The multiple and essential roles of LUBAC in inflammation and immune signaling have been extensively investigated. A large number of studies show that dysregulation of LUBAC is associated with autoinflammation and immunodeficiencies.

The canonical NF-κB pathway is well known to mediate key pro-inflammatory gene expression, playing an important role in both adaptive and innate immunity (Lawrence, 2009; Hoesel and Schmid, 2013). LUBAC is now known for its close involvement in the activation of the NF-κB pathway induced by various stimuli such as TNF, PGN, CD40-L, LPS, IL-1β etc (Ikeda et al., 2011; Gerlach et al., 2011; Damgaard et al., 2012, 2013). The synthesis and covalent attachment of linear ubiquitin chains to key elements in the upstream signaling pathways of NF-κB are a prerequisite for NF-κB activation. These important upstream pathways are related to both immune response and apoptosis, such as the TNFR-signaling pathway (Reviewed in detail by Rittinger and Ikeda, 2017 and Spit et al., 2019).

The TNFR-signaling pathway is one of the most well-characterized pathways in NF-κB activation (Figure 1.17). In this pathway, the binding of a proinflammatory cytokine TNFα to the transmembrane TNF receptor (TNFR) initiates the trimerization of TNFR. Then a group of signaling cofactors

including TRADD, TRAF2, RIP1 and cIAP1/2 are recruited to the trimeric TNFR to form the “TNFR complex I” (Barnhart and Peter, 2003). E3 ligase cIAP1/2 mediates the K63-polyubiquitination of the protein kinase RIP1, which then allows the recruitment of the kinase TAK1/TAB complex to RIP1. The phosphorylation of the IKK complex by TAK1/TAB is required for IKK activation. cIAP1/2 also auto-ubiquitinates itself with K63-linked ubiquitin chains, which is essential for recruiting LUBAC. LUBAC subsequently targets linear ubiquitin chains to NEMO, one of the subunits in the IKK complex (Haas et al., 2009; Gyrd-Hansen and Meler, 2010). The UBAN domain of NEMO shows high affinity to M1 linkages and thus facilitates the activation of kinase IKKα/β, another two components of the IKK complex, to phosphorylate IκBα (Rahighi et al., 2009). IκBα is an inhibitor to the heterodimeric transcriptional factor NF-κB. Once IκBα is phosphorylated, it is recognized by the E3 ligase SCF-βTrCP for K48-mediated proteasomal degradation, therefore releasing the inhibitory state of NF-κB. The liberated NF-κB transcriptional factor is then translocated into the nucleus, allowing the gene expression of pro-inflammatory factors, such as cytokines, chemokines, and adhesion molecules (Lawrence, 2009; Liu and Chen, 2011; Pamukcu et al., 2011).

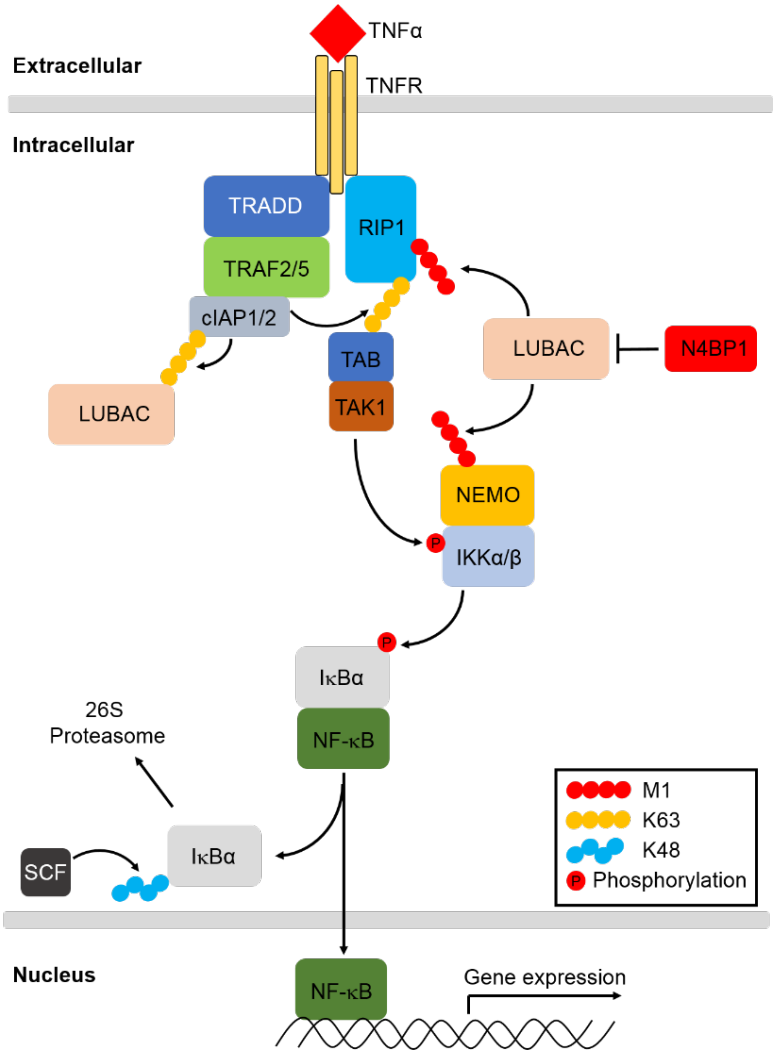


Figure 1.17 Schematic representation of NF-κB activation by the TNFR signaling pathway.

Similarly, in PGN-induced NOD2 immune signaling, E3 ligase XIAP ubiquitinates RIP2 and recruits LUBAC for the downstream activation of NF-κB (Damgaard et al., 2012; Damgaard et al., 2013).

LUBAC is fine-tuned by multiple factors among which N4BP1 (NEDD4-binding protein 1) was newly identified as a novel regulator by negatively regulating LUBAC activity (Kliza et al., in preparation). The mechanism will be discussed in the result chapters of this thesis.

1.3 Ubiquitin E3 ligases

1.3.1 Classification of ubiquitin E3 ligase families

The ligation between ubiquitin and the substrate catalyzed by E3 ligase is the final step of ubiquitination. A large number of E3 ligases have been identified and the diversity enables their involvements in targeting specific substrates. E3 ligases can be classified into 3 main families, the RING-, HECT- and RBR-type E3 ligases (Figure 1.18) (Morreale and Walden, 2016).

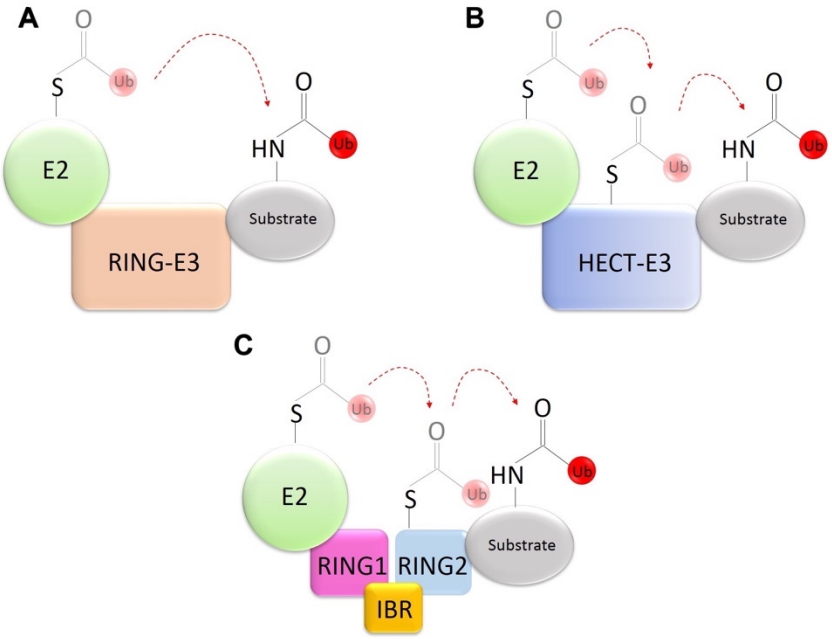


Figure 1.18 Schematic representation of the ubiquitin transfer mechanisms by the three ubiquitin E3 ligase families. A. RING-type E3. B. HECT-type E3. C. RBR-type E3. Ubiquitin moieties that link to E2 or E3 via thioester are shown in pale red. Ubiquitin moieties that link to substrates via isopeptide bond are shown in solid red. Dashed arrow indicates the ubiquitin transfer process.

1.3.1.1 RING (*Really Interesting New Gene*)-type E3 ligases

The RING-type E3 family has more than 600 members. It is characterized by containing either the RING domain or the U-box domain. Both domains bind to E2 and share similar fold. As shown in Figure 1.19, the RING domain has a unique cross-brace motif where two zinc ions are

coordinated (Metzger et al., 2014). The U-box domain does not coordinate Zn, instead the fold is maintained by hydrogen bonds and salt bridges (Ohi et al., 2003).

RING-type E3s bind to both substrate and ubiquitin-charged E2 (Ub~E2), thus acting as a scaffold to facilitate the direct ubiquitin transfer from Ub~E2 to the substrate. Therefore, the aminolysis reaction that results in the formation of isopeptide bond between ubiquitin and the substrate is conducted by E2 (Figure 1.18).

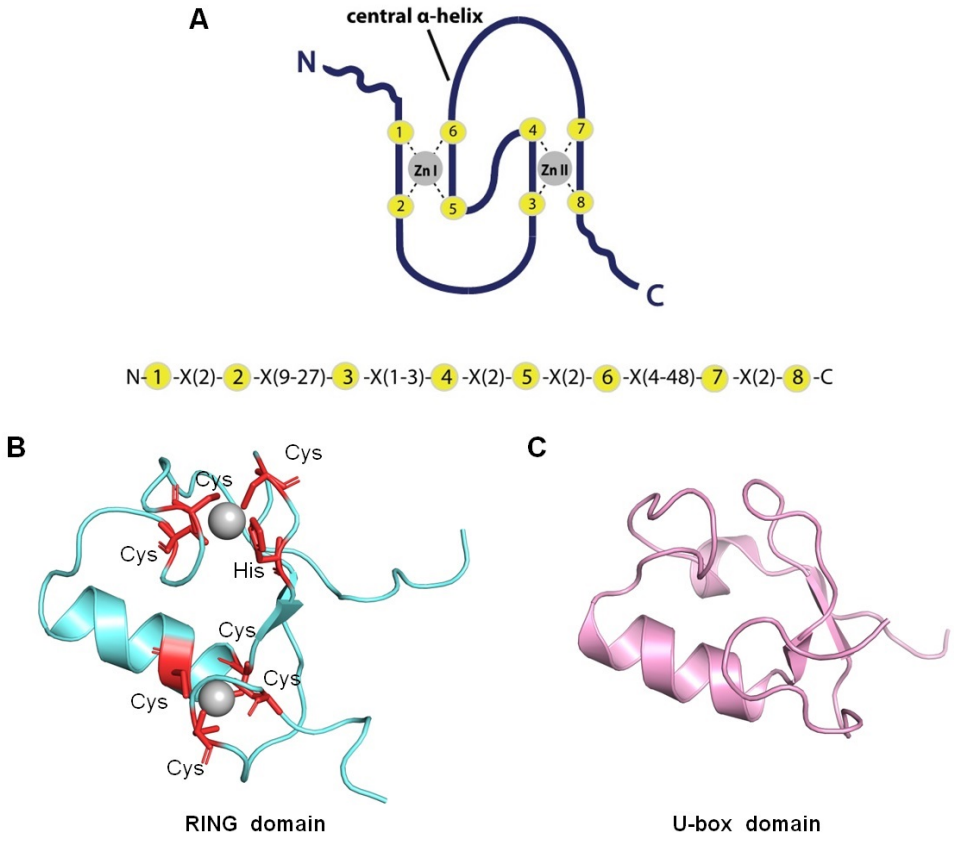


Figure 1.19 Structure of the RING/U-box domain. A. Schematic of the RING domain organization for Zn-coordination (© 2014 Metzger et al., by permission, Licence No: 4882050067905). The residues that coordinate Zn are shown in yellow with number 1-8. X represents any residue. B. Crystal structure of the RING domain from c-Cbl (PDB: 1FBV). The residues that coordinate Zn are indicated in red. C. Solution structure of the U-box domain from Prp19 (PDB: 1N87). Figures were generated by Pymol according to the respective PDB structure entry.

As mentioned in section 1.1.2.1, the Cullin protein family could interact with some RING-type E3 ligases to form the Cullin/RING ubiquitin ligases (CRLs). CRLs are highly diversified due to multiple protein combinations. Taking the SCF complex (Skp, Cullin, F-box containing complex) as an example, it plays key roles in the regulation of cell cycle by targeting proteins for 26S proteasomal degradation. The SCF complex is composed of four subunits among which Cullin serves as a scaffold to organize RBX1, Skp1 and Skp2. RBX1 is the RING-type E3 ligase for E2-binding while the F-box protein Skp2 is responsible for substrate recognition. Skp1 is the adaptor protein that links Skp2 to Cullin (Figure 1.20).

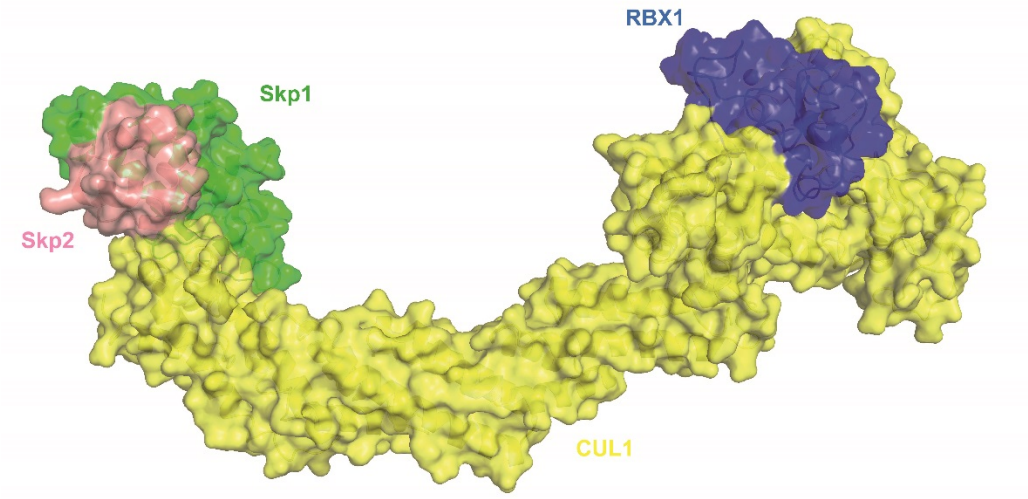


Figure 1.20 Structure of the SCF complex. The complex is shown in the surface mode. Figure was generated by Pymol using PDB: 1LDK.

The interaction between the dimerized RING domain from RNF4 and the stable E2-Ub conjugate provides some insights into the activation of the ubiquitin-charged E2 by the RING-type E3 ligase (Figure 1.21, right panel). To help form the crystal, a stable UbcH5A-Ub conjugate is used where the catalytic C85 of UbcH5A is mutated to lysine to form a stable isopeptide bond with G76 of Ub, in order to mimic the thioester between C85 and G76. As seen from the complex structure (PDB: 4AP4), each RNF4 RING domain binds to one UbcH5A-Ub conjugate while each ubiquitin contacts both RING

moieties from the dimer. This induces the conjugated ubiquitin folded back onto UbcH5A, in comparison to the structure where ubiquitin non-covalently binds to the E2 (Figure 1.21, left panel). The interaction extends the contact between the ubiquitin C-terminus and the E2 active site, which thus facilitates the nucleophilic attack from the ubiquitin C-terminal glycine to the incoming substrate (Plechanovová et al., 2012).

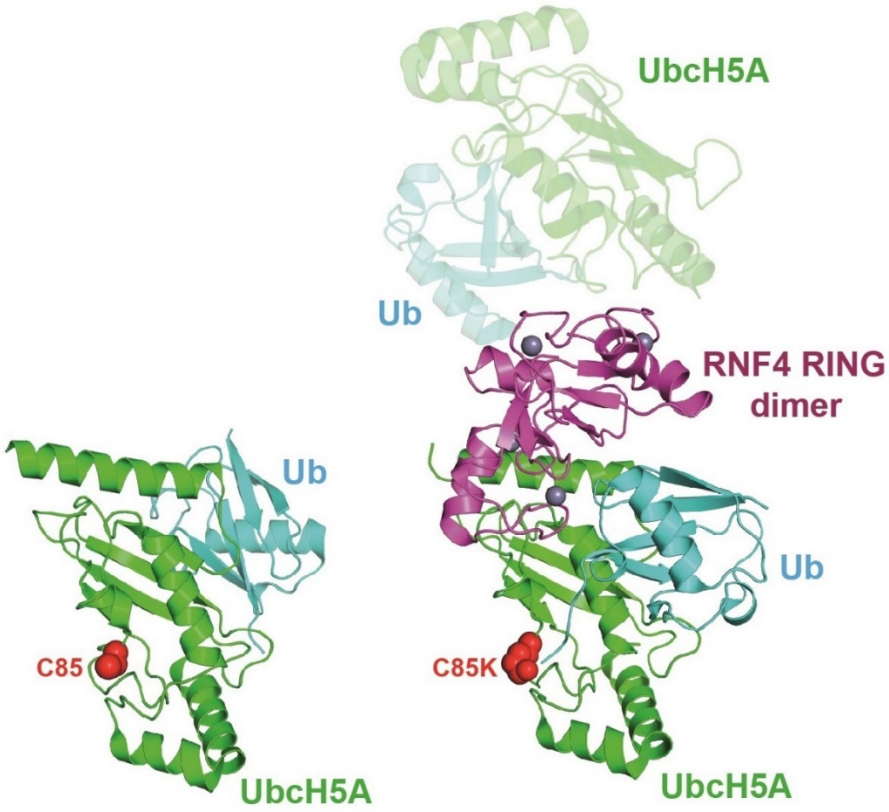


Figure 1.21 Structure of the RING/U-box domain. Left panel. Structure of UbcH5A non-covalently bound to ubiquitin (PDB: 3PTF). The catalytic C85 of UbcH5A is indicated in a red sphere. Right panel. Structure of the dimerized RNF4 RING domain in complex with two stable UbcH5A-Ub conjugates (PDB: 4AP4) where the C85 is mutated to lysine (indicated in a red sphere) to form a non-hydrolysable isopeptide bond with G76 of Ub. The UbcH5A/ubiquitin complexes from both structures are shown in the same position for a better comparison. Figure were generated by Pymol according to the respective PDB structure entry.

1.3.1.2 HECT (*H*omologous to *E*6AP *C*arboxyl *T*erminus)-type E3 ligases

The HECT-type E3 family features the HECT domain, which is composed of the N terminal lobe and the C terminal lobe with a flexible linker in between the two lobes. HECT domain utilizes the N lobe to recruit Ub~E2 while the catalytic Cys residue from the C lobe accepts the ubiquitin from E2 via the transthiolation reaction to form the Ub~E3. Finally, the ubiquitin charged E3 catalyzes the aminolysis reaction by transferring ubiquitin to the substrate to form the stable isopeptide bond (Figure 1.18).

Early structural studies of the E2 enzyme UbcH7 in complex with the HECT domain from E6AP (PDB: 1C4Z) shows that the distance between the catalytic Cys from both molecules is ~ 40 Å in the absence of ubiquitin, which is far from achieving efficient transthiolation (Figure 1.22A) (Huang et al., 1999). Later, the crystal structure of Ub~UbcH5B-HECT^{NEDD4L} complex was resolved (PDB: 3JW0), giving new insights into the ubiquitin transfer mechanism (Kamadurai et al., 2009). In this structure, the ubiquitin-loaded E2 brings its active Cys in close proximity at less than 8 Å to that of HECT. It indicates that drastic conformational changes upon binding facilitate the ubiquitin transfer from E2 to E3 (Figure 1.22B).

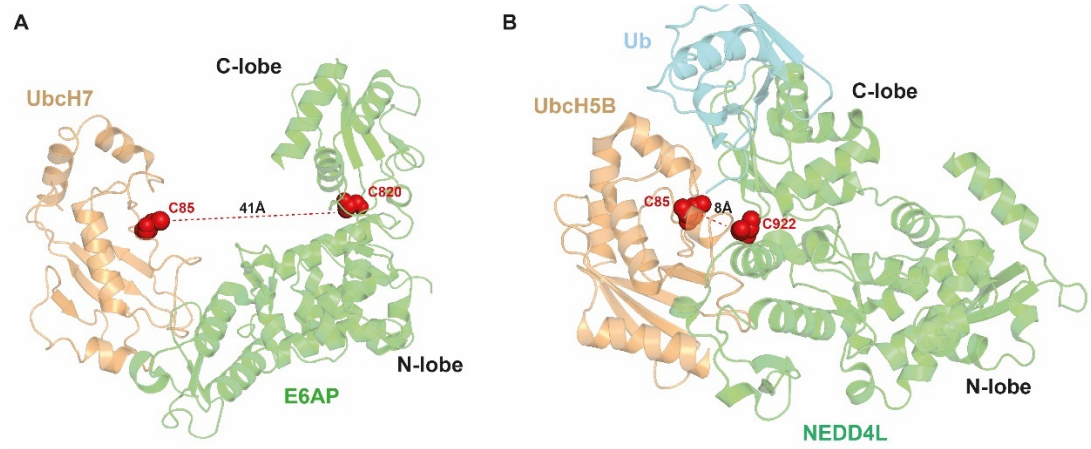


Figure 1.22 Structure of the HECT domain in complex with E2. A. Structure of UbcH7-HECT^{E6AP} complex (PDB: 1C4Z). B. Structure of Ub~UbcH5B-HECT^{NEDD4L} complex (PDB: 3JW0). The two E2s are placed in a similar position for a better comparison. The

distance between the catalytic Cys from E2 and E3 is indicated by the dashed lines. Figure were generated by Pymol according to the respective PDB structure entry.

1.3.1.3 RBR (*R*ING-in-*b*etween-*R*ING)-type E3 ligases

The RBR-type E3 ligases are discovered much later. It consists at least 14 members in human and yet the function from most of them is still unclear. Among them, the most studied proteins include Parkin, HHARI (also known as ARIH1), HOIP and HOIL-1L (Figure 1.23A).

Parkin mediates a variety of ubiquitination processes including mono-ubiquitination, K6, K11, K48- and K63-linked polyubiquitination, to be associated with mitochondrial autophagy (mitophagy), Parkinson's disease etc (Chung et al., 2001; Vives-Bauza et al., 2010; Cunningham et al., 2015). Independently, Parkin was also found to act as a transcriptional suppressor to regulate p53 expression by interacting with p53 promoter (Da Costa et al., 2009). HHARI has been reported to participate in cellular proliferation, DNA damage response, neuronal differentiation etc (Aguilera et al., 2000; Elmehdawi et al., 2013; von Stechow et al., 2015). HOIP and HOIL-1L are components of LUBAC, playing key roles in innate immunity and inflammatory reactions (See section 1.1.2.3).

1) The RBR domain

RBR-type E3 ligases are more like a RING/HECT hybrid. As the name suggests, the RING1 and RING2 domain are separated by the IBR (in-between-RING) domain. All the three domains in RBR coordinate Zn ions. Although RBR ligases contain a RING domain, the catalytic mechanism of this E3 is in an HECT-like fashion. The RBR^{RING1} domain, which adopts a similar structure to the canonical RING domain in a cross-brace fold, recognizes and recruits E2~Ub. While the catalytic Cys is localized in the RBR^{RING2} domain to accept the activated ubiquitin from E2. The RBR^{RING2} domain is structurally different from the canonical RING domain, but similar

to the RBR^{IBR} domain (Figure 1.23B, C) (Morreale and Walden, 2016; Dove and Klevit, 2017).

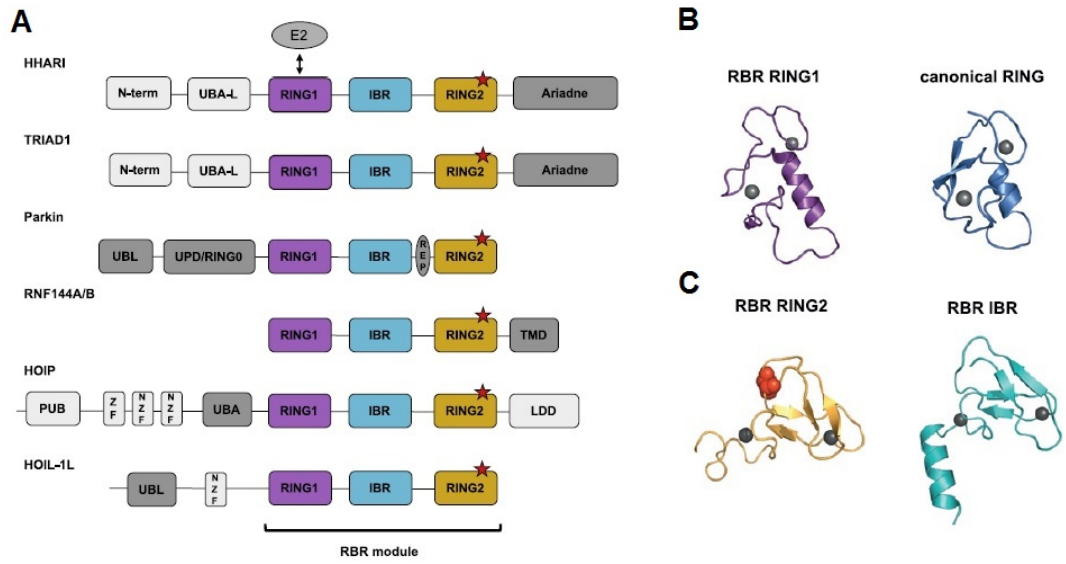


Figure 1.23 An overall view of the RBR-type E3 ligases (© 2017 Dove and Klevit, by permission, Licence No: 4882040764839). A. Domain structures of representative RBR E3s. Red stars indicate the active Cys site. B. The RBR^{RING1} domain from HHARI (PDB: 5UDH) is structurally similar to the canonical RING domain from BRCA1 (PDB: 1JM7). C. The RBR^{RING2} domain adopts similar fold to the RBR^{IBR} domain (both domains are from HHARI, PDB: 5UDH). Dark gray circles in B and C indicate Zn ions.

Though the similar structure between the RBR^{RING1} domain and the canonical RING domain, the conformation state of E2~Ub upon binding to RBR- or RING-type E3 is very different. As exemplified in Figure 1.21, E2~Ub adopts a closed state when attached to the canonical RING domain, increasing the reactivity of E2 toward the Lys sidechain amino group from the substrate. In this closed conformation, ubiquitin is in contact with E2, resulting in the burial of its hydrophobic patch. With the help from RING-type E3 ligases, ubiquitin is then transferred directly from E2 to the substrate. In contrast, in the case of HHARI, HOIP and RNF144A, E2~Ub is in an open state upon binding to the RBR^{RING1} domain while the ubiquitin

hydrophobic patch is exposed, facilitating the transthiolation process from E2~Ub to E3~Ub (See the example from Figure 1.26B,C,D) (Dove and Klevit, 2017).

2) Auto-inhibition of RBR-type E3s

Auto-inhibition is a common feature within the RBR-type E3 family. However, the mechanisms of auto-inhibition between RBR-type E3s are quite diverse due to their various domain organizations. Correspondingly, the release of the auto-inhibition is organized in different ways.

Two examples for autoinhibition from HHARI and Parkin are shown in Figure 1.24. In the apo form of HHARI, the catalytic Cys from the RING2 domain is partially buried in the interface between RING2 and the Ariadne domain. RING1 is far away from RING2 at a distance of ~ 90 Å. Note that though upon E2~Ub binding, HHARI is still in its auto-inhibited state. The catalytic center of HHARI is ~ 54 Å away from that of E2 (UbcH7), suggesting an unsuitable conformation for catalysis at this state (Duda et al., 2013; Yuan et al., 2017). As for Parkin, the UPD/RING0 domain buries the catalytic Cys from RING2 at the interface. REP, the linker between IBR and RING2 obstructs E2 binding to RING1 (Wauer and Komander, 2013; Trempe et al., 2013). The auto-inhibition implicates that appropriate conformation rearrangement is required for RBR-type E3s to induce full activity before the ubiquitin transfer.

The release of the auto-inhibitory state varies, and most was discovered by functional studies. For instance, the RBR ligase activity of HHARI is activated by binding to the neddylated Cullin-RING ligase (CRL) complexes (Kelsall et al., 2013). Parkin phosphorylation by kinase PINK1 is required in order to release its auto-inhibition when responses to mitochondrial damage (Eiyama et al., 2015; Pickrell et al., 2015). The UBA domain from HOIP auto-inhibits its catalytic RBR domain (Stieglitz et al., 2012). The UBL domain from HOIL-1L or SHARPIN can both release the auto-inhibition of

HOIP (Discussed in more details in section 1.3.3.1) (Gerlach et al., 2011; Ikeda et al., 2011; Kirisako et al., 2006; Tokunaga et al., 2009, 2011).

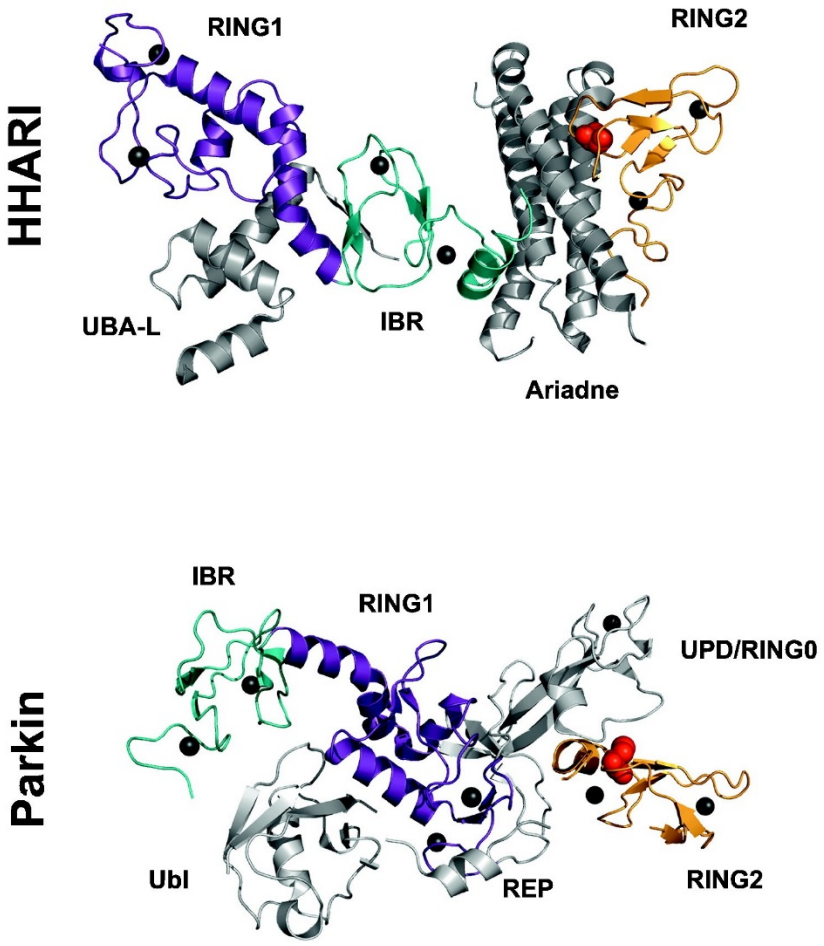


Figure 1.24 The structure of HHARI (PDB: 4KBL) and Parkin (PDB: 4K95) in their respective auto-inhibition state (© 2017 Dove and Klevit, by permission, Licence No: 4882040764839). The catalytic Cys in each RING2 domain is indicated as a red sphere while Zn ions are indicated as dark circles.

1.3.2 HOIP

Apart from the above general introduction of the RBR-type E3 ligases, as the study of the ligase activity of HOIP is one of the topics in this thesis, a short review of HOIP is discussed in the following section.

1.3.2.1 Auto-inhibition of HOIP

As mentioned in section 1.1.2.3, HOIP plays the main catalytic role in LUBAC. Similar to other RBR members, the catalytic RBR domain from the apo HOIP is auto-inhibited by its UBA domain (Stieglitz et al., 2012). The structure of auto-inhibited HOIP is still unavailable. Current studies reveal that the UBL domains from both HOIL-1L and SHARPIN share high sequence and structure similarity (Figure 1.25A). They can both bind to the UBA domain of HOIP, thus releasing HOIP from auto-inhibition. Binding of HOIL-1L and SHARPIN is therefore required to transform HOIP into a catalytic competent state (Gerlach et al., 2011; Ikeda et al., 2011; Kirisako et al., 2006; Tokunaga et al., 2009, 2011).

Despite the high similarity, the two UBL domains interact with $HOIP^{UBA}$ in different manners, revealed by the structures of corresponding complexes (Yagi et al., 2012; Liu et al., 2017). $SHARPIN^{UBL}$ interacts with the N-terminal $HOIP^{UBA}$, while HOIL-$1L^{UBL}$ contacts the C-terminus of $HOIP^{UBA}$ (Figure 1.25B). Furthermore, adding both $SHARPIN^{UBL}$ and $HOIP^{UBA}$ together to HOIP synergistically activates HOIP, resulting in a stronger catalytic capacity to synthesize linear ubiquitin chains compared to the interaction with only one of the two proteins (Liu et al., 2017). Recently, the crystal structure of the trimeric complex $HOIP^{UBA}$/$SHARPIN^{UBL}$/HOIL-$1L^{UBL}$ has been solved, which elucidates how the UBA domain of HOIP interacts with both UBL domains simultaneously (Fujita et al., 2018). The LUBAC-tethering motifs (LTMs) from both $SHARPIN^{UBL}$ and HOIL-$1L^{UBL}$ form a single globular domain, which provides extra supports to stabilize the LUBAC trimeric complex (Figure 1.25C).

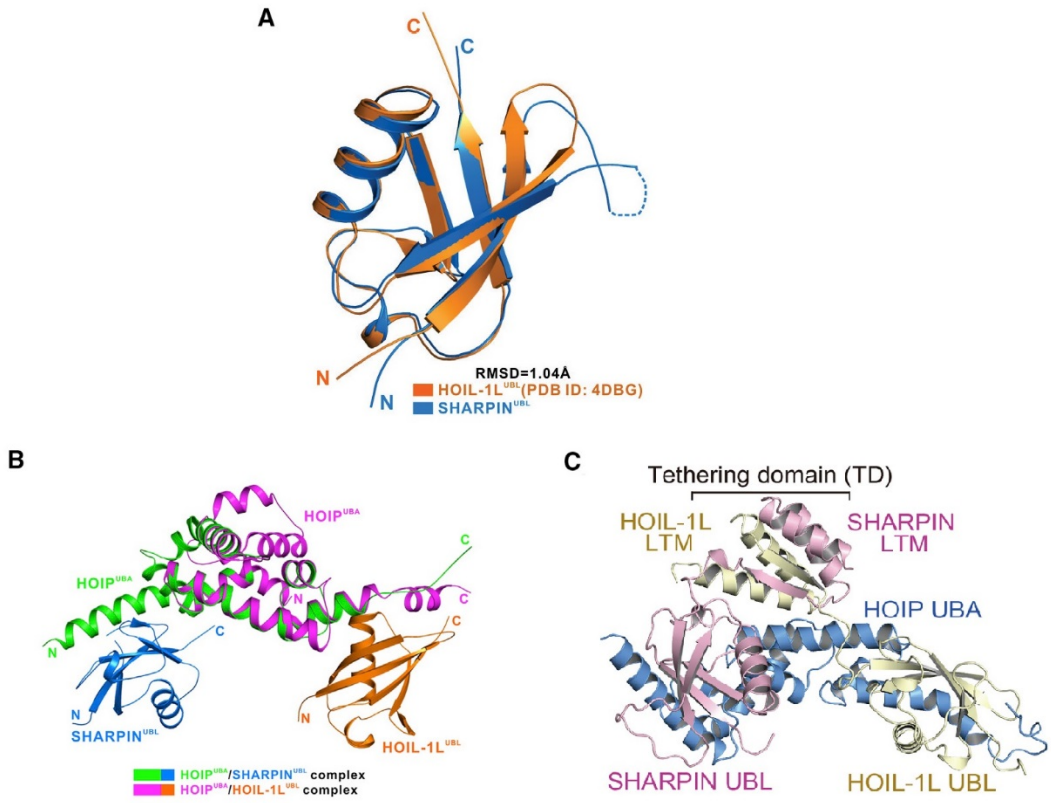


Figure 1.25 Structure insight into the release of HOIP auto-inhibition by the UBL domain from SHARPIN and HOIL-1L. A. HOIL-1L^{UBL} and SHARPIN^{UBL} shares high similarity in structure (© 2017 Liu et al., by permission). B. HOIL-1L^{UBL} and SHARPIN^{UBL} binds to different areas in HOIP^{UBA} (© 2017 Liu et al., by permission). C. The LUBAC-tethering motifs (LTMs) from SHARPIN and HOIL-1L form the TD domain in the trimeric LUBAC core to stabilize LUBAC complex (PDB: 5Y3T) (© 2018 Fujita et al., by permission).

1.3.2.2 Catalytic mechanism of HOIP

Similar to the HECT-type E3 ligase, the RBR-type E3 ligase itself determines the linkage type of the synthesized ubiquitin chains. In the case of the isolated HOIP^{RBR} domain (including the C-terminal extension), no ubiquitin chains are formed when N-terminal His-tagged mono-ubiquitin is used as the substrate, indicating that the Met1 from the ubiquitin is involved in the chain formation. In the *in vitro* ubiquitin assay, the HOIP^{RBR} domain could accept the ubiquitin from E2~Ub and then form a thioester between

HOIP^{RBR} and ubiquitin (RBR~Ub). Cys885 in the RING2 domain is identified as the key catalytic residue responsible for the ligase activity (Stieglitz et al., 2012).

The structure of the HOIP RING2L domain (the RING2 domain including the C-terminal LDD domain) in complex with ubiquitins has been solved at 1.6 Å resolution (PDB: 4LJO) (Figure 1.26A) (Stieglitz et al., 2013). The LDD domain together with the zinc-finger forms a ubiquitin-binding platform to non-covalently interact with the acceptor ubiquitin (Ub_{acc}). It specifically positions the α-amino group of Met1 from the acceptor ubiquitin to the HOIP catalytic core. The thioester is formed between the C-terminal Gly76 carboxyl group of the donor ubiquitin (Ub_{don}) and the Cys885 thiol group from HOIP ^{RING2}. The α-amino group of the first residue Met1 in the acceptor ubiquitin is protonated by His887, the catalytic base from HOIP, therefore the acceptor ubiquitin is capable of being the nucleophile reagent to attack the above thioester. The reaction results in the simple peptide bond formation between the donor and acceptor ubiquitin. Being the critical catalytic base, His887 is pH-dependent. The H887A mutant can restore its function in higher pH conditions.

The crystal structure of HOIP^{RBR} with UBcH5B~Ub transfer complex has been solved to further provide structural evidence to the ubiquitin transfer process from E2 to HOIP (PDB: 5EDV) (Lechtenberg et al., 2016). In a single asymmetric unit, the RING1-IBR domain from one HOIP molecule and the RING2L domain from another HOIP molecule forms a “clamp-like” structure, to engage the E2~Ub (Figure 1.26B, D).

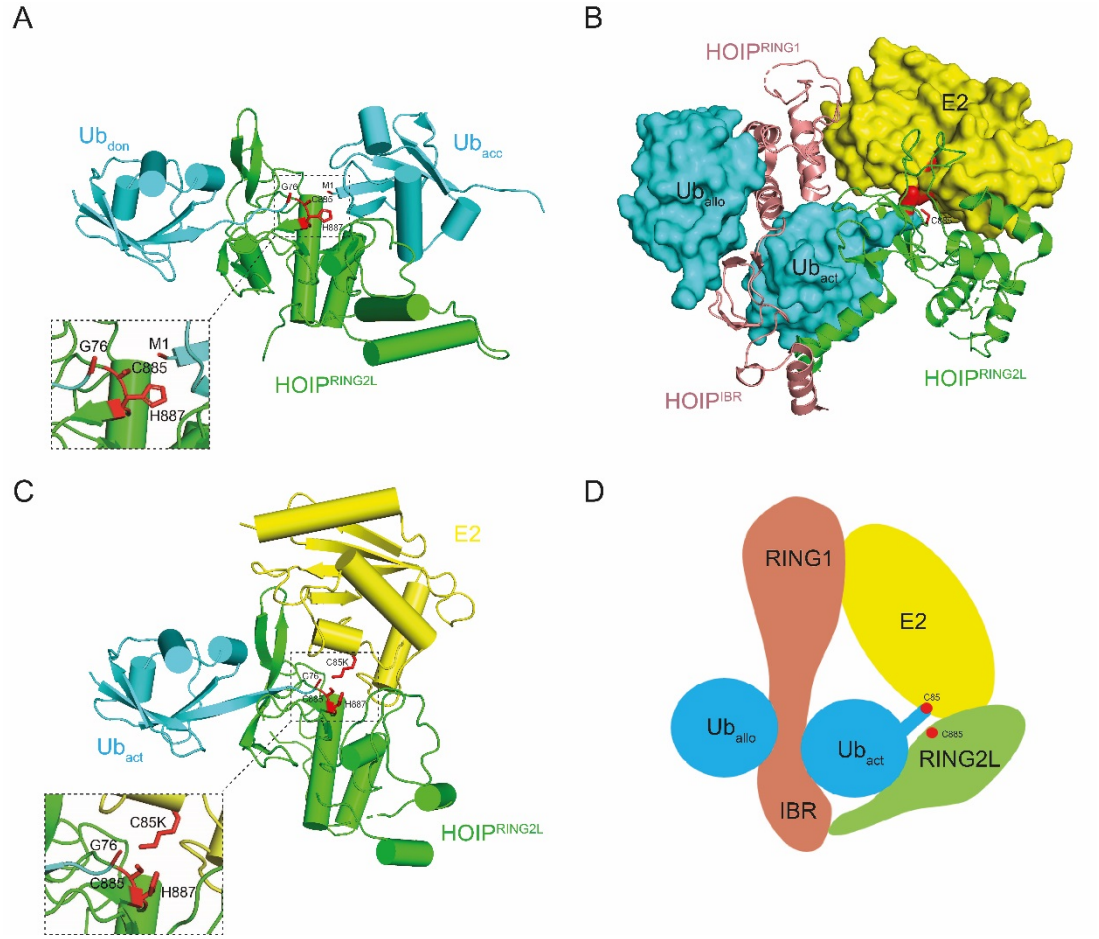


Figure 1.26 Structure insight of the catalytic mechanism of HOIP. A. The crystal structure of the HOIPRING2L/Ub·Ub transfer complex (PDB: 4LJO). Ubdon: the donor ubiquitin. Ubacc: the acceptor ubiquitin. G76 from the donor ubiquitin, M1 from the acceptor ubiquitin and the catalytic C885, H887 from HOIP are highlighted. B and C. The crystal structure of the HOIPRBR/E2~Ub transfer complex (PDB: 5EDV). Uballo: the allosteric ubiquitin. Ubact: the activated ubiquitin from E2~Ub. Uballo and RING1-IBR are not shown in Figure C. To mimic the thioester bond between E2 and Ub, the catalytic C85 from E2 (UBcH5B) was mutated to lysine in order to form a stable isopeptide bond with G76 from ubiquitin. G76 from the activated ubiquitin, C85K from E2 and the catalytic C885, H887 from HOIP are highlighted. D. Schematic illustration of the structure from Figure B. Both RING2L domains from Figure A and C are placed in the same position for a better comparison. Figures were generated by Pymol according to the respective PDB structure entry.

The mechanism underlying the activation of HOIP RBR domain and the following ubiquitin transfer from E2 to HOIP has been proposed (Lechtenberg et al., 2016). First of all, the auto-inhibition state of HOIP is released by HOIL-1L, enabling ubiquitin (or linear chains) binding to the RBR domain. The binding of a ubiquitin (Ub_{allo}) to the RING1 domain allosterically rearranges the conformation of the RING1-IBR arm, facilitating the engagement of E2~Ub onto RING1 and RING2L. The interaction between the activated ubiquitin from E2~Ub and the RING2L domain further stabilizes the E2~Ub engagement. Therefore, the E2~Ub is clamped into the HOIP RBR domain in an appropriate position which brings the active Cys from HOIP in close proximity to the E2~Ub thioester, thus enables the ubiquitin transfer to HOIP.

It is worth noticing that when comparing the structure of the $HOIP^{RING2L}$/Ub·Ub transfer complex (PDB: 4LJO, Figure 1.26A) with this $HOIP^{RBR}$/E2~Ub transfer complex (Figure 1.26C), the acceptor ubiquitin-binding regions in RING2L is overlapped with that from E2, which indicates that the two transfer states of HOIP are mutually exclusive. In other words, once finishing the delivery of the activated ubiquitin to HOIP, E2 must exit from the complex to spare room for the acceptor ubiquitin in order to achieve the subsequent linear chain formation catalyzed by HOIP. Similarly, to keep this turnover, the growing chain or the ubiquitinated substrate has to later vacate from HOIP, thus enables the loading of a new E2~Ub.

1.4 An introduction to the RNAse-containing proteins

The functional study of N4BP1 (NEDD4 binding protein 1) as an inhibitor on LUBAC ligase activity and as a linear ubiquitin receptor is the major topic of this thesis. Initially, from a phylogenetic study, N4BP1 has been implicated as an RNAse due to the NYN (*N*4BP1, *Y*acP-like *N*uclease) domain it contains (Figure 1.27) (Anantharaman and Aravind. 2006),

In the NYN phylogenetic tree, there are several proteins whose domains are structurally close to N4BP1. It's worth noting that NYN domain in these proteins is fused to several other domains related to ubiquitin binding such as the CUE domain, the UBA domain etc, indicating the possibility involved in the ubiquitin-related processes. Among them, some structural studies are available for KIAA0323 (also called KHNYN) and MCPIP1 (also known as Regnase-1 or Zc3h12a).

The nomenclature of the NYN domain reported in different literature is very promiscuous. For example, the NYN domain from MCPIP1 was named alternatively as the NCD or NCR or PIN domain from different studies. The NYN domain from N4BP1 is usually named as the RNAse domain. To avoid confusion, due to that the NYN domain functions as an RNAse, it will be uniformly named as the RNAse domain for N4BP1, KHNYN and MCPIP1 in this thesis hereafter. In this section, the three representative RNAse-containing proteins will be briefly introduced.

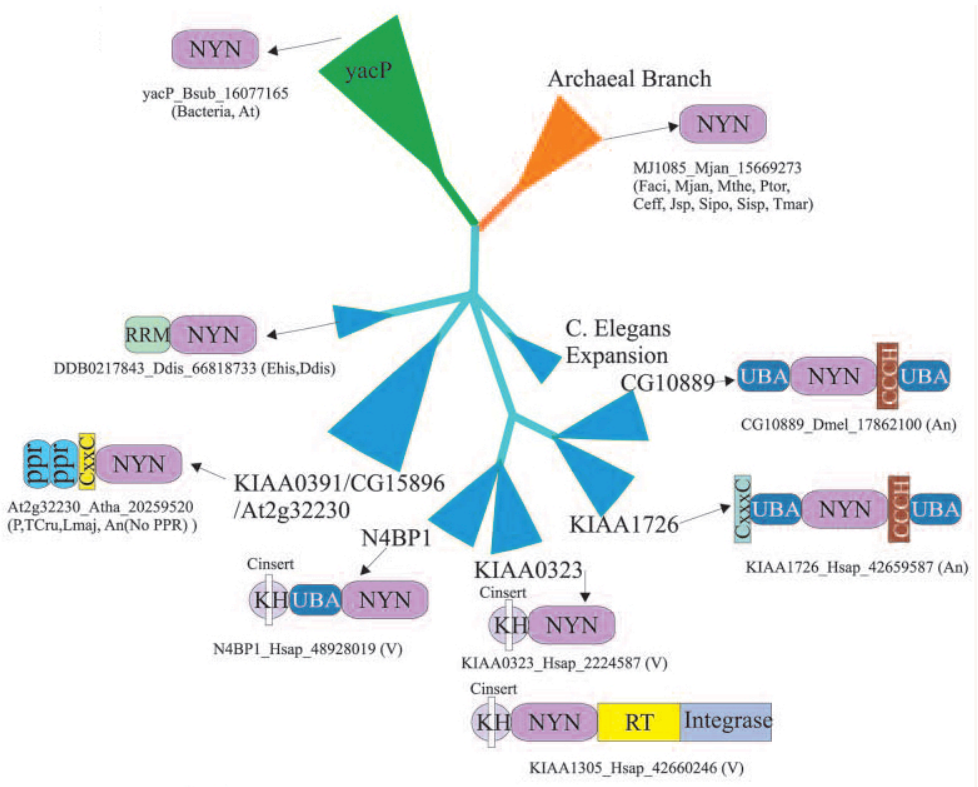


Figure 1.27 Phylogenetic trees of the proteins containing the RNAse (NYN) domain (© 2006 Anantharaman and Aravind, by permission). Note that KIAA0323 is KHNYN. Though MCPIP1 is not shown in the tree, another MCPIP family member (KIAA1726 or MCPIP3) is present. The NYN domain will be called as RNAse domain hereafter in this thesis.

1.4.1 N4BP1 (NEDD4 binding protein 1)

N4BP1, as the name suggests, was first identified to interact with Nedd4, a HECT-type E3 ligase by yeast two-hybridization. *In vivo* experiments have demonstrated that N4BP1 can be mono-ubiquitinated by Nedd4 (Murillas et al., 2002). Further study has shown that N4BP1 can be also poly-ubiquitinated by Nedd4 and then undergoes proteasomal degradation (Sharma et al., 2010). Interestingly, sumoylation on N4BP1 was identified in the same study and this modification negatively regulates its poly-ubiquitination. The sumoylated N4BP1 cannot be poly-ubiquitinated and thus its stability is maintained.

N4BP1 also interacts with another HECT-type E3 ligase ITCH, but not a substrate of ITCH. Instead, N4BP1 binding to the ITCH WW2 domain reduces the affinity of ITCH to other substrates (eg. p73, p63 and c-Jun) in a competitive manner, thus diminishing the poly-ubiquitination of these substrates (Oberst et al., 2007).

Several binding partners of N4BP1 have been identified by pull-down experiments combined with mass spectrometry in the wild-type neuroblastoma. Among them, CEZANNE, a deubiquitinating enzyme involved in the NF-κB pathway is significantly enriched. CEZAANE removes the ubiquitination modification on TRAF3, thus helping TRAF3 escape from being degraded. The recruitment of CEZANNE by N4BP1 results in the stabilization of TRAF3, suggesting that N4BP1 plays as a negative regulator in the NF-κB pathway (Spel et al., 2018).

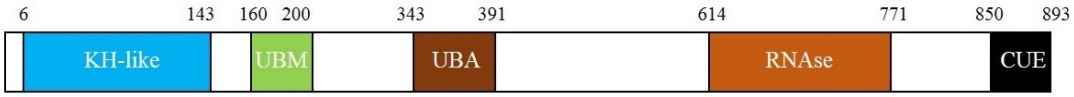


Figure 1.28 The domain structure of the full-length mouse N4BP1 studied in this thesis. KH: K Homology. UBA: Ubiquitin association. UBM: Ubiquitin binding motif. CUE: coupling of ubiquitin conjugation to ER degradation domain.

Recently, the RNAse activity of N4BP1 has been demonstrated. The expression of N4BP1 can be induced by type-I interferons (IFNs), which suggests a role in defending the infection from viruses. In primary T cells and macrophages, N4BP1 acts as a suppressor to HIV-1 on account of its RNAse activity, while D623 is identified as the key catalytic residue for the activity (Yamasoba et al., 2019). It is also reported that N4BP1 is negatively regulated by MALT1. MALT1 cleaves N4BP1 at its R509 site, which results in the inactivation of N4BP1, thus further promoting the reactivation of latent HIV-1 (Yamasoba et al., 2019).

Apart from the above findings, N4BP1 was recently discovered as a linear ubiquitin receptor from a yeast-two hybrid (Y2H) screening aiming for identifying linear ubiquitin binding domains (LUBIDs) by our collaborators (K. Kliza, K. Husnjak et al.,). From the cell biological studies, N4BP1 is recruited to the TNFR complex upon TNFα stimulation. Cell lines which are deficient for N4BP1 showed significantly increased levels of TNFα-induced NF-κB transcription activity as well as the accumulation of linear ubiquitinated substrates. The C-terminus of N4BP1 recognizes linear ubiquitin chains with high specificity from *in vitro* pull-down experiments, but the mechanism behind the ubiquitin selectivity is still unknown. It is also found that the N4BP1 N-terminus interacts with HOIP, the LUBAC catalytic subunit that mediates linear ubiquitin chain synthesis. Therefore, N4BP1 is proposed to be a negative regulator of the TNFR signaling pathway possibly by regulating LUBAC activity, but the mechanism requires in-depth investigation. Interestingly, N4BP1 can be cleaved by CASP8 into N- and C-terminal fragments during apoptosis. The N-terminus of N4BP1 could still negatively regulate the TNFR signaling pathway while the C-terminal cleavage fragment undergoes proteasomal degradation (Kliza et al., unpublished data).

1.4.2 KHNYN (KH and NYN domain-containing protein)

KHNYN was initially described in the phylogenetic study of the NYN domain (as shown in Figure 1.27) and predicted to have the RNAse activity (Anantharaman and Aravind. 2006). So far, KHNYN is still poorly characterized.

Recently, KHNYN is reported to interact with the zinc finger antiviral protein (ZAP). ZAP is responsible to recognize and bind to CpG dinucleotide region in viral RNA but lacks the nuclease activity. KHNYN is thus recruited and functions as an RNAse to degrade viral RNA (Ficarelli et al., 2019).

KHNYN is also found to be associated with NEDD8. KHNYN binds to neddylated cullins (including CUL1 to CUL4) and is independent on ubiquitination. Unmodified cullins lost the binding ability to KHNYN, indicating the direct interaction between NEDD8 and KHNYN. Later, the C-terminal domain from KHNYN was identified to be responsible for the binding. Therefore, this domain is named as CUBAN (Cullin-Binding domain Associating with NEDD8) (Castagnoli et al., 2019).

CUBAN domain shows significant binding affinity to NEDD8 rather than ubiquitin. It is worth noticing that N4BP1 was also screened from the same research by phage lambda display, but N4BP1 interacts with ubiquitin instead of NEDD8. The binding behaviour is completely opposite to that of KHNYN, indicating the different binding modes between the two proteins. CUBAN domain alone (PDB: 2N5M) adopts a three-helix bundle arrangement which is commonly shared with other UBDs. However, the location of the second helix is in an opposite plane, making CUBAN structurally distinguished from UBDs such as the CUE domain or the UBA domain (Figure 1.29A).

The co-structure of the KHNYN C-terminus CUBAN domain with NEDD8 has been solved by solution NMR (PDB: 2N7K). Polar and hydrophobic interactions are mainly involved in the interaction between the two molecules. In addition, the higher chemical shift perturbations (CSPs) of the residues in the loop1 area from CUBAN suggest spatial rearrangement upon binding to NEDD8. Although NEDD8 I44A mutant has completely abolished the binding with CUBAN, the co-structure shows no direct contact between the NEDD8 I44 hydrophobic patch and CUBAN (Figure 1.29B) (Castagnoli et al., 2019). It was suggested that the strong CSP changes of those indirect residues (including I44) are coming from the “far effect” that was propagated from the intramolecular disturbance when the binding event happens (Santonico et al., 2019).

Additional dynamic properties were analysed by NMR titrations at different CUBAN: NEDD8 ratio. At the early stage where there is only insufficient NEDD8 added to CUBAN, the hydrophobic interaction plays the dominant role in the binding. While increasing NEDD8 to the same ratio to CUBAN, the perturbations were extended to larger surface areas with the involvement from several polar and neutral residues. Although the perturbed CUBAN surfaces are quite similar upon either NEDD8 or ubiquitin at the final 1:1 ratio, the intermediate evolvement seems quite limited in the case of CUBAN/ubiquitin where only a few hydrophobic residues are perturbed at the early binding stage (Santonico et al., 2019).

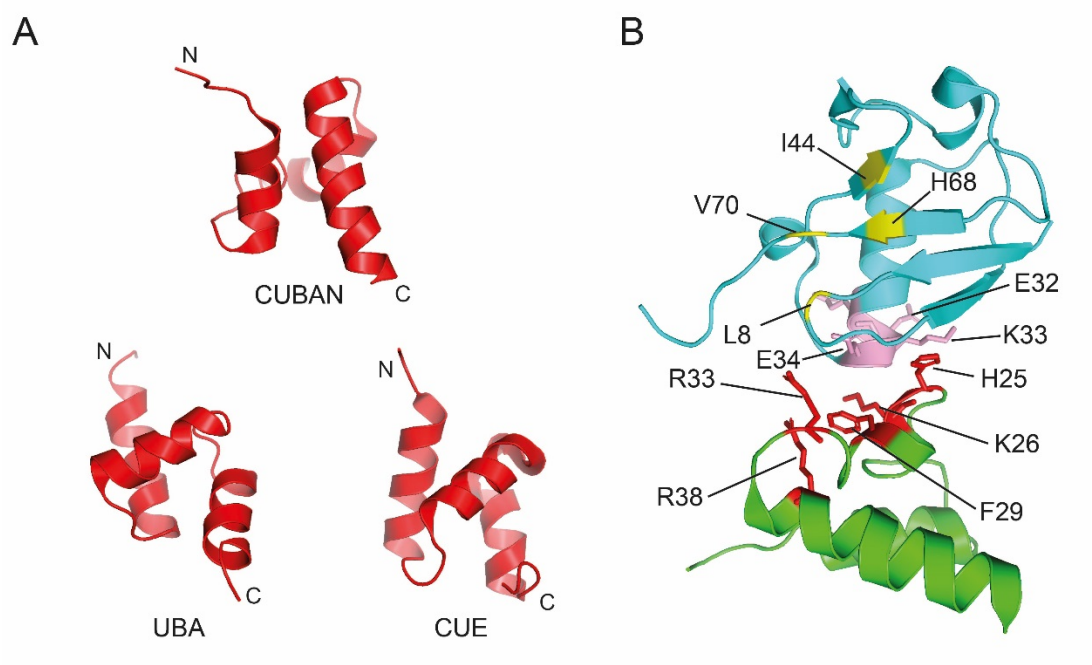


Figure 1.29 The structure of KHNYN^{CUBAN} domain as well as the complex with NEDD8. A. Overall structure comparison among the KHNYN CUBAN domain (PDB: 2N5M), the gp78 CUE domain (PDB: 2EJS) and the UBA domain from the NSFL1 cofactor p47 (PDB:1V92). B. The solution NMR structure of the CUBAN/NEDD8 complex (PDB: 2N7K). The residues involved in the interaction are shown in sticks and highlighted in pink (from CUBAN) and red (from NEDD8) respectively. The residues from the I44 hydrophobic patch of NEDD8 are coloured in yellow. Figures were generated by Pymol according to the respective PDB structure entry.

1.4.3 MCPIP1 (MCP-1-induced protein 1)

MCP-1-induced protein (MCPIP) family, as the name suggests, was first found being highly induced by MCP-1 in human peripheral blood monocytes (Zhou et al., 2006). The MCPIP members feature a novel CCCH-zinc finger motif, among which MCPIP1 is significantly induced by LPS in the macrophage activation process. MCPIP1 overexpression inhibits the promoter activity of TNFα when induced by LPS or p65, suggesting a negative role in regulating NF-κB activation in macrophages (Liang et al., 2008). MCPIP1 functionally suppresses T and B cell activation. Deficiency of MCPIP1 in mice results in developing autoimmune diseases. MCPIP1 was also found to be the substrate of MALT1 under T cell receptor (TCR) stimulation. The cleavage at R111 site by MALT1 releases the inactivation states of T cells (Uehata et al., 2013).

The RNAse activity of MCPIP1 has been identified. It degrades cytokine mRNA such as *IL-6*, *IL-12β* by targeting their 3'-UTR regions (Matsushita et al., 2009; Xu et al., 2012). The crystal structure of the MCPIP1 RNAse domain (PDB: 3V33) displays a conserved catalytic centre when superimposed with the PIN domain from SMG6 (PDB: 2HWW) (Figure 1.30A, B). The catalytic centre is located within a negatively-charged pocket which consists of several conserved acidic residues. Among them, the single mutations of D141, D225, D226 and D244 result in complete loss of the RNAse activity (Figure 1.30C) (Xu et al., 2012).

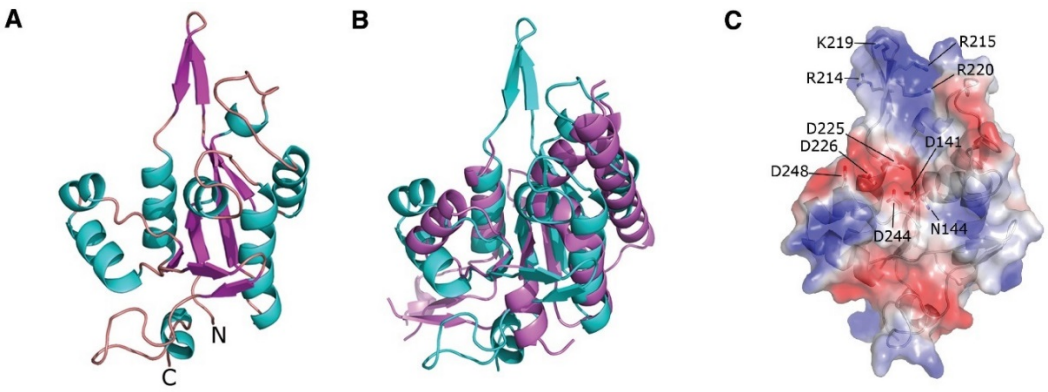


Figure 1.30 Crystal structure of the MCPIP1 RNAse domain (© 2012 Xu et al., by permission, Licence No: 4881950673039). A. Overall structure of the RNAse domain (PDB: 3V33). B. Superimposition of the MCPIP1^{RNAse} domain with the SMG6^{PIN} domain (PDB: 2HWW). C. Electrostatic surface of MCPIP1^{RNAse} domain. The catalytic center featuring several conserved acidic residues (D141, N144, D225, D226, D244 and D248) are located in the negatively-charged pocket. Positively-charged residues are indicated in blue while negatively-charged residues are indicated in red.

Further study shows that the head-to-tail dimerization of the RNAse domain is critical to exert the RNAse activity of MCPIP1 *in vitro*. Several single point mutations on the dimer interface break the dimerization and have shown no activity towards *IL-6* mRNA (Figure 1.31A). The N-terminal domain (NTD) was found to help increase the RNAse activity. The direct interaction between the NTD domain and the RNAse domain was revealed by NMR. According to the above structural characterization, the regulation of the RNAse activity of MCPIP1 has been proposed: The NTD domain binding to the RNAse domain stops the formation of the head-to-tail RNAse oligomer, thus only the RNAse dimer is functioning to carry out the RNAse activity (Figure 1.31B) (Yokogawa et al., 2016).

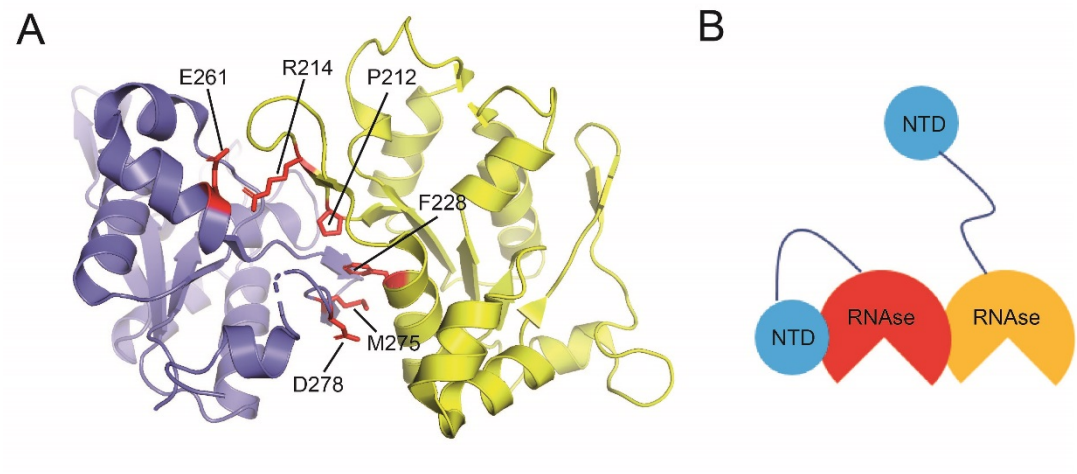


Figure 1.31 Oligomerization of MCPIP1^{RNAse} domain is critical for its RNAse activity. A. Dimer structure of the MCPIP1 RNAse domain (PDB: 5H9V). The two RNAse molecules are coloured in yellow and blue respectively. Key residues that maintain the dimer states are highlighted in red sticks. B. Proposed working model of the regulation of MCPIP1 RNAse activity by domain-domain interaction and dimerization.

Apart from the regulation on the mRNA level, surprisingly, the deubiquitinating activity from the same RNAse domain in MCPIP1 was also discovered. Sequence analysis reveals that MCPIP1^{RNAse} domain contains both the Cys box and Asp box, which are usually characterized in Cys protease. The sequence alignments of the Cys box between MCPIP1 and UCHL1 shares 39% homology and the catalytic Cys is conserved. MCPIP1 could cleave K48- and K63-linked polyubiquitin chains *in vitro*. Furthermore, MCPIP1 is able to remove the ubiquitin modification on TRAFs (including TRAF2, TRAF3 and TRAF6), thus negatively regulates JNK and NF-κB activity. This finding demonstrates its crucial roles in inflammatory signaling pathways on the protein level (Liang et al., 2010).

Because of its newly identified deubiquitinating activity, the MCPIP family is then grouped into the 6th member of DUB family (Fraile et al., 2012). It is interesting to investigate the possible relations between the dual catalytic roles of MCPIP1 as a negative regulator in the inflammatory signaling.

1.5 Objective of this thesis

The study of N4BP1 is the central topic of this thesis. As mentioned in section 1.4.1, N4BP1 is identified as a novel linear ubiquitin receptor that negatively regulates the TNFα signaling pathway. It is hypothesized that the role of N4BP1 being a negative regulator is mediated by inhibiting the E3 ligase activity of LUBAC core subunit HOIP as well as substrate recruitment under pro-inflammatory conditions.

Based on the above knowledge and hypothesis, the study in this thesis will characterize both in depth the inhibition module and the specific ubiquitin-interacting module from N4BP1 by combining multiple approaches including biochemistry, biophysics and structural biology (Figure 1.32). This study will further expand our understanding of the physiological functions of N4BP1 in inflammatory and innate immunity.

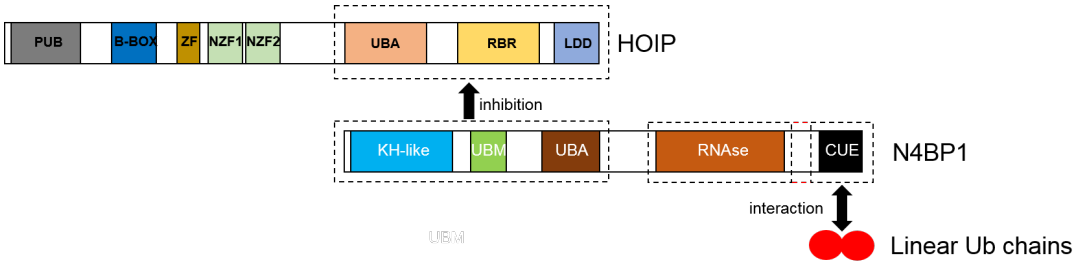


Figure 1.32 Schematic representation of the possible involvements of N4BP1 in regulating HOIP and ubiquitin binding.

CHAPTER II

NMR theory

2.1 Introduction

Nuclear magnetic resonance (NMR) spectroscopy is one of the most powerful and modern techniques to characterize the molecular structures ranging from small organic chemicals, polymers to large biomolecules (polypeptides or proteins). In addition, NMR is also applied to study chemical reactions, molecular dynamics etc. Thus, it has been widely used in areas including chemical, material, biological and medical sciences. There is another important branch from NMR application, called magnetic resonance imaging (MRI), which has been developed into an essential and indispensable technique for medical diagnosis.

The NMR phenomenon was first discovered by American physicist Isidor Isaac Rabi. He was awarded the Nobel Prize in Physics in 1944. Then the development and application of NMR have highlighted several key moments in the Nobel Prize history. The development of the methods for NMR measurements helped two American scientist Felix Bloch and Edward Mills Purcell win the Nobel Prize in Physics in 1952. The applications of high-resolution NMR spectroscopy in chemistry were significantly developed by Swiss scientist Richard Robert Ernst, who was awarded the Nobel Prize in Chemistry in 1991. The contributions to the development of solution NMR for determining biomacromolecules by Swiss scientist Kurt Wüthrich helped him win the Nobel Prize in Chemistry in 2002 (Boesch, 2004; Shampo et al., 2012).

In the structural biology field, there are many advantages for applying NMR spectroscopy to study protein structures. For example, the protein is measured in solution conditions, which are close to its native physiological states. Compared to crystallography, sample preparation for biomolecular NMR does not require the time-consuming crystal screening procedure. NMR is also an ideal approach to study the dynamics of protein folding, domain motions, in particularly protein-ligand interactions or

protein-protein interactions, providing rich information on the binding behaviour, such as disassociation constant (Kd), perturbed surface etc.

Needless to say, NMR also has its limits in characterizing biomolecules. The proteins are usually uniformly labelled with isotopes (^{15}N or ^{13}C or both) by recombinant expression. Considering the high cost of the isotopes, it requires the protein to be expressed at a reasonable level in minimal media. The protein is also required to be stable and properly folded. The protein size is usually limited in NMR study due to the increasing difficulties to resolve the crowded signals from a complex spectrum. Small proteins (<20kDa) are ideal for NMR, though there are cases where the structures of large proteins were successfully solved, it requires advanced spectrometers equipped with more powerful magnets. As much more signals acquired from a large protein, analyzing the complex multi-dimensional spectra can be challenging.

2.2 Basic NMR theory

Nuclear magnetic resonance (NMR) is the phenomenon where a nucleus in a magnetic field transits between discrete energy levels at a certain radiation frequency when exposed to electromagnetic radiation. NMR uses radio frequency (RF) radiation. A nucleus with a spin quantum number $I \neq 0$ is NMR-active, thus NMR has selectivity towards certain nuclear types.

2.2.1 Nucleus energy splitting and resonance

The nucleus is positively charged. For a nucleus with a non-zero spin quantum number, its spin generates a magnetic field. Therefore, it has a magnetic moment μ and can be considered as a very tiny magnetic bar (Figure 2.1A). When no external magnetic field is applied, nuclei are spinning at random orientations. However, when the external magnetic field (referred as "B_o" hereafter and usually conventionally placed along the z-axis) is present, the nuclei will either align with the magnetic field or align against the magnetic field B_o, thus displaying different spin states with different energy levels (Figure 2.1C).

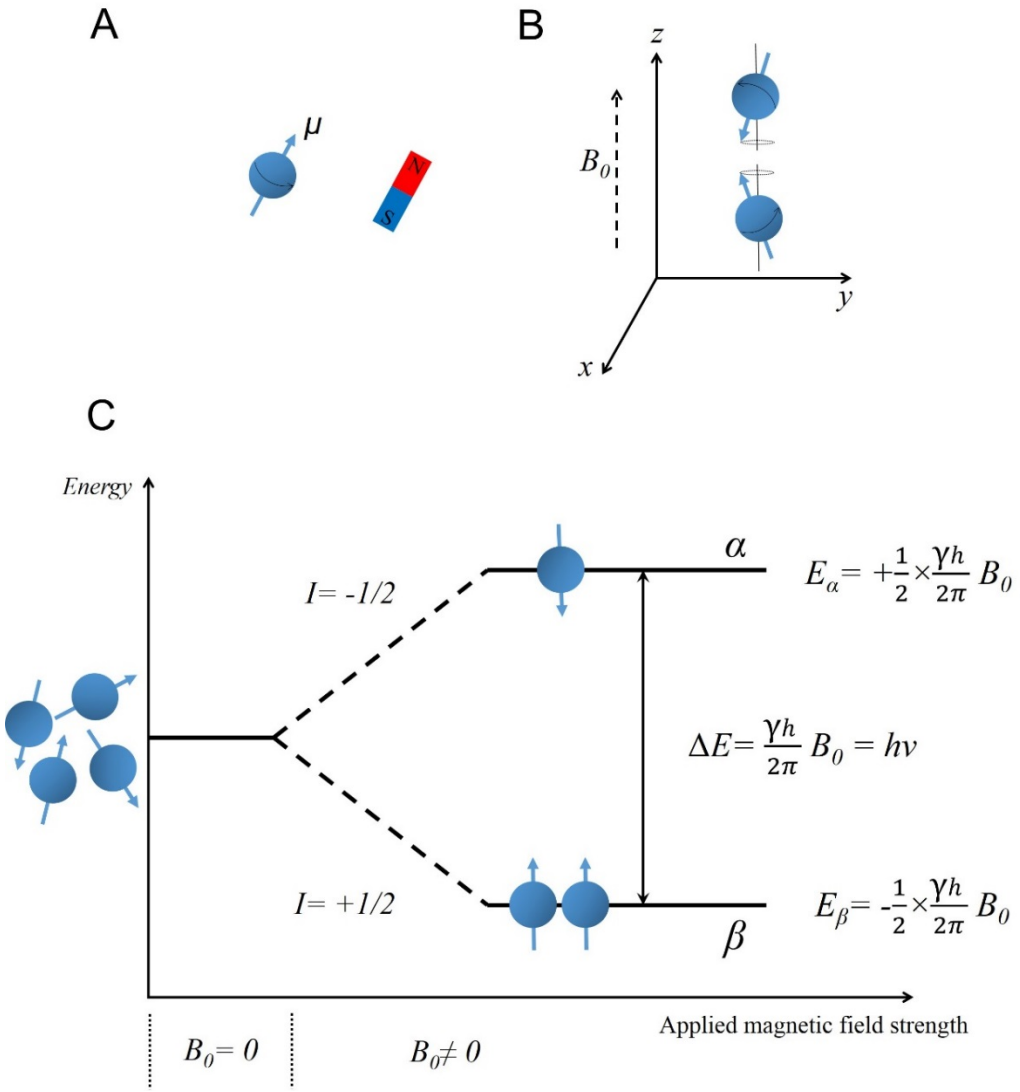


Figure 2.1 Schematic representations of the nucleus energy splitting (I=1/2) in an applied magnetic field. A. A nucleus with a spin quantum number I≠0 has the property to spin and thus has a magnetic moment μ. It can be considered as a very tiny magnetic bar. B. Larmor precession of nuclei in the applied magnetic field B_o. B_o is usually conventionally applied along the z-axis. C. Energy splitting of the nuclei (I=1/2) when the external magnetic field B_o is applied.

We also need to know that the nucleus does not completely align parallel or antiparallel with the applied magnetic field. The nucleus will also precess about the magnetic field with an angle, similar to the movement of a spinning

top. This is called Larmor precession and its characteristic frequency is thus called Larmor frequency ν (Figure 2.1B). The Larmor frequency is unique to each nucleus.

The most studied nuclei for NMR include ^{1}H, ^{13}C and ^{15}N etc. The spin quantum numbers of these nuclei are all at 1/2. The number of spin states in an applied magnetic field is thus 2I+1=2. Taking proton (^{1}H) as an example, in an applied external magnetic field B_o, proton energy splits into 2 sublevels and thus protons display two spin states and populate into two species (Figure 2.1C). The two populations follow the Boltzmann distribution. There is only a tiny excess of the number of the lower energy species than the higher energy species. From a macroscale view, it would result in the bulk magnetization of a group of nuclei aligning parallel with B_o at equilibrium.

To excite the protons from the lower energy state to the higher energy state, the RF radiation at the proton Larmor frequency ν must be applied to exactly fulfil the energy gap ΔE between the two states (Equation 1). Nuclei can only absorb the energy from RF radiation that is equal to ΔE in order to resonate. The energy transition during this spin flip can be recorded in an NMR spectrum.

$$\Delta E = \frac{\gamma h}{2\pi} B_o = h\nu \quad (1)$$

$$\nu = \frac{\gamma}{2\pi} B_o \quad (2)$$

Where:

h is the Plank's constant (6.626×10^{-34} Js)

γ is the gyromagnetic ratio of the nucleus

B_o is the applied magnetic field

ν is the Larmor frequency of a certain nucleus

2.2.2 Chemical shift

The discussion from above is the simplest case where the proton is considered as an isolated particle. However, in real fact, a nucleus is

surrounded by various chemical environments, mostly from electrons. In an applied external magnetic field B_o, electron spins opposite to the nucleus. Hence, a small magnetic field B_e from the electron is induced, opposite to B_o (Figure 2.2). This results in a reduced total external magnetic field B_{eff} for the nucleus (Equation 3). This is called the shielding effect.

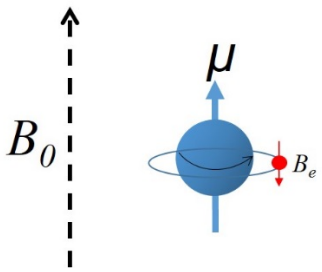


Figure 2.2 Schematic representation of the shielding effect from the electron to the nucleus. Electron is shown as a small red ball. The red arrow indicates the magnetic field B_e generated by the spinning electron.

ρ is the shielding constant in the below equation. Due to the shielding effect, the Larmor frequency of a nucleus will be changed, which is described by the chemical shift. Even for the same nucleus, the different surrounding electron distributions would result in different chemical shifts, providing important information on its local chemical environments. Thus, it has become a key parameter for structure determination.

$$B_{eff}=B_o-B_e=(1-\rho)\,B_o \qquad (3)$$

Replacing B_o with B_{eff} in Equation 1 will result in an increased value of Bo in order to fulfil the resonance.

$$B_o = \frac{2\nu\pi}{(1-\rho)\gamma} \qquad (4)$$

According to Equation 4, the chemical shift (represented by the Larmor frequency ν here) is proportional to the external magnetic field B_o. It will result in various chemical shift values if the same nucleus is placed in NMR spectrometers with different magnetic field strengths. To avoid this inconvenience, chemical shift δ is usually represented in relative to a

standard chemical (usually using TMS) in the below form (Equation 5). Thus, chemical shift δ is a ratio without unit. It is usually expressed in ppm.

$$\delta = \frac{\delta sample - \delta ref}{\delta ref} \quad (5)$$

Where:

δ_{sample} is the absolute resonance frequency of the sample

δ_{ref} is the absolute resonance frequency of the standard

2.2.3 *J*-Coupling

J-coupling (also called spin-spin coupling or scalar coupling) is the indirect interaction between two magnetically active nuclei, through chemical bonds. Considering a nucleus (I≠0) as a small magnetic bar, its magnetic moment could affect a neighbouring nucleus (I≠0) and *vice versa*. This mutually leads to the energy splitting for both nuclei in the magnetic field, resulting in the splitting of the NMR signals (Figure 2.3).

The strength of the coupling is expressed as the constant *J*, which is the frequency difference between the split signals. *J* constant is the intrinsic property of the nucleus and is independent of the external magnetic field. *J*-coupling usually occurs as far as three bonds. Thus, it provides distance and dihedral angle information of two coupled nuclei.

J-coupling is the basic principle of some key multidimensional NMR experiments, for example, double resonance HSQC and triple resonance approaches such as HNCA, HN(CO)CA etc. These experiments are based on the magnetization transferred between certain types of nuclei via J coupling, providing selective nuclear correlations that are important for protein assignments.

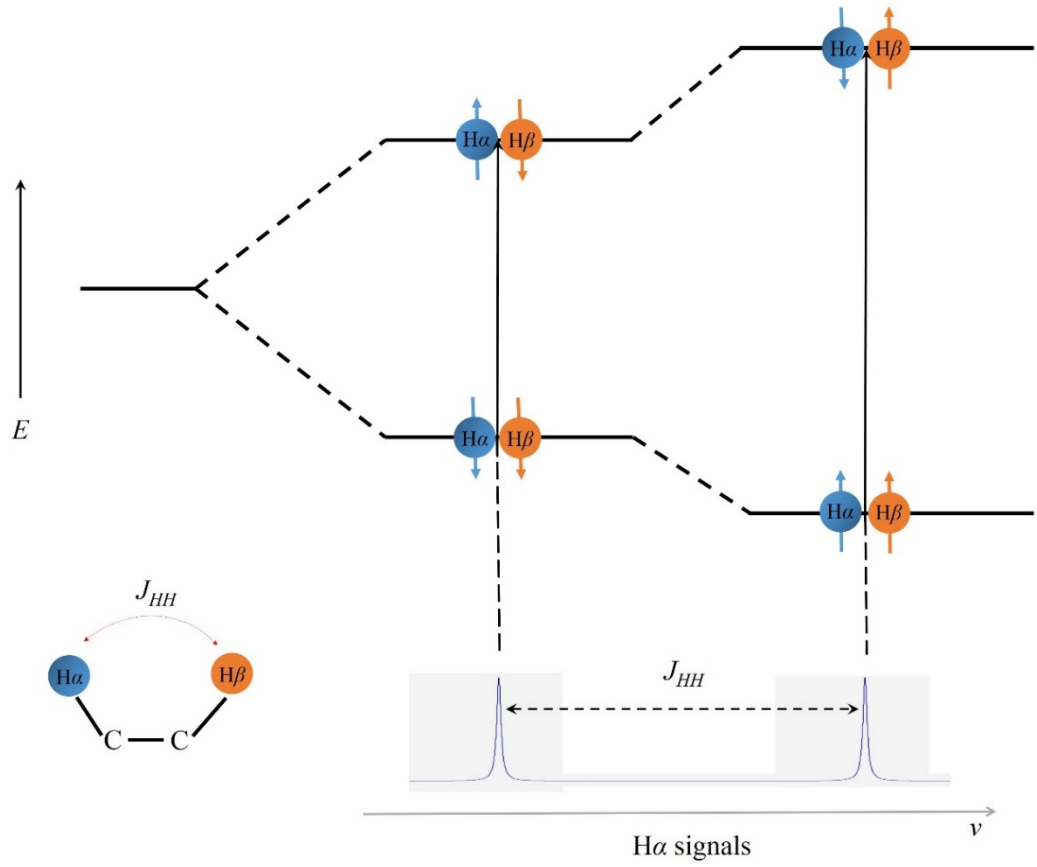


Figure 2.3 Schematic representation of an example of *J*-coupling between two protons. The energy levels of Hα are affected by the spin states from a nearby Hβ, as described as *J*-coupling, resulting in the splitting of the Hα NMR signals. J_{HH} is the *J*-coupling constant.

2.2.4 The Nuclear Overhauser Effect (NOE)

Different from the *J*-coupling where the nuclei interact indirectly through bonds, two spin-active nuclei also have the direct interaction through space. This is described as dipolar coupling. Key factors involved in dipolar coupling include the distance between two nuclei as well as their orientations etc. Therefore, the dipolar coupling provides information on the geometric structure of the molecule.

Based on dipolar coupling, the NOE effect is discovered and has been developed into several essential applications for NMR. The NOE effect is

described as the changes of the signal strength from an observed nucleus when a nearby nucleus (usually within 5 Å) is excited or saturated by radio frequency. The NOE strength η is directly correlated to the distance d between two nuclei:

$$\eta \propto 1/d^6 \tag{5}$$

The NOE effect is particularly important for the NOESY (Nuclear Overhauser Effect Spectroscopy) experiments. In the simplest two-dimensional H-H NOESY experiment (Figure 2.4), magnetization is exchanged between all the protons that have the NOE effects. That is to say, any protons close enough to a certain proton of interests are visible in the NOESY spectrum. It provides rich distance information on the intramolecular or even intermolecular structure.

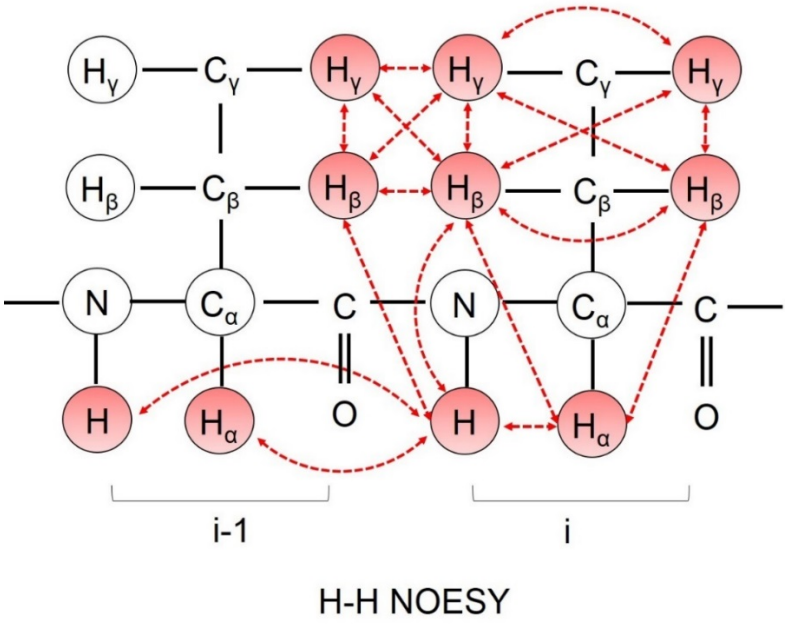


Figure 2.4 Schematic representation of the interactions between a group of close protons detected in H-H NOESY. Arrows indicate the magnetization transfer between protons that have NOE effects. Atom signals visible in the NOESY spectrum are colored in red.

2.3 One-dimensional ^{1}H NMR

One-dimensional ^{1}H NMR experiment is the simplest and most commonly used experiment in the starting stage of sample characterization. The natural abundance of ^{1}H makes it the most sensitive element for NMR detection. In a modern Fourier-Transform (FT) NMR spectroscope, a single-pulse sequence is applied to obtain the 1D ^{1}H NMR spectrum (Figure 2.5). After the 90° RF pulse is applied, the bulk magnetization of the protons flips to the x-y plane compared to its original direction (along the z-axis). The signals of magnetization are recorded in the relaxation process where the bulk magnetization is returning to the z-axis over this period. This process is called Free Induction Decay (FID). The FID signals are further transformed from the time domain to the frequency domain by Fourier transformation in order to generate the conventional spectrum that is usually seen from an NMR spectroscope.

1D ^{1}H NMR spectrum of a small organic compound has only a few signals and is easy to resolve due to the simplicity of the structures. However, when it comes to even a small protein, a large number of signals are overlapped, thus the complexity of the spectrum is greatly increased and difficult to solve. To collect as much structural information as possible, multi-dimensional heteronuclear NMR spectroscopy has been developed, including double and triple resonance experiments.

However, it is still useful to roughly evaluate the protein quality by 1D NMR, for example, how well the protein is folded. It is helpful to verify the feasibility to collect more complicated multidimensional NMR spectra.

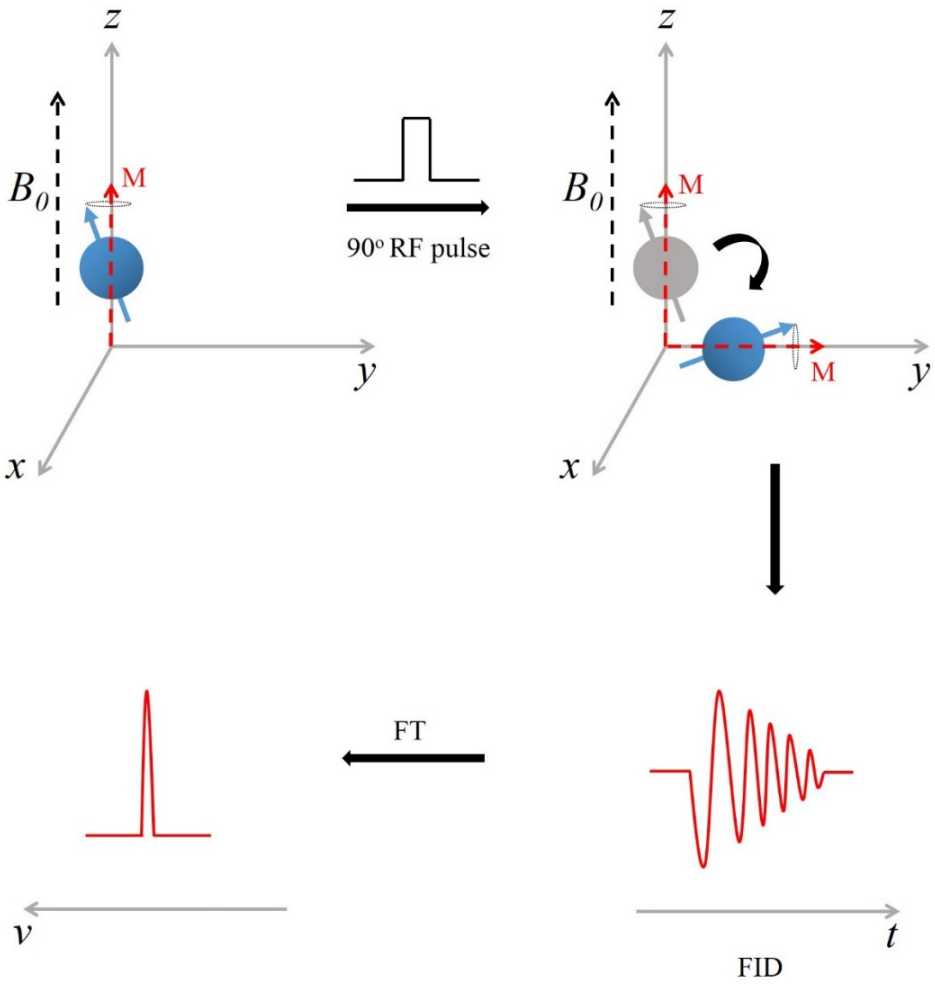


Figure 2.5 Schematic representations of the process of 1D NMR experiment applied with a single 90° pulse. A 90° RF pulse was applied to the nuclei to induce the bulk magnetization (M) flipping to the x-y plane. The Free Induction Decay (FID) signals over time in the magnetization relaxation process are recorded and then Fourier-transferred to the frequency domain.

2.4 two-dimensional HSQC

HSQC (Heteronuclear Single Quantum Correlation) experiment is the double resonance approach to detect the correlation between two different nuclear species. As a two-dimensional NMR experiment, the chemical shift of ^{1}H is usually recorded in the direct dimension, while ^{15}N or ^{13}C is in the indirect dimension.

$^{1}H/^{15}N$-HSQC is perhaps the most widely used 2D NMR spectrum which collects all the H-N correlations via *J*-coupling, including the backbone amide groups and those from the side chains of Trp, Asn and Gln. Therefore, each signal in $^{1}H/^{15}N$-HSQC can be assigned to a certain residue (except those from the side chains). Due to fewer numbers of NH groups than CH groups of a protein, $^{1}H/^{15}N$-HSQC spectrum is less crowded than $^{1}H/^{13}C$-HSQC. Together with its high sensitivity, $^{1}H/^{15}N$-HSQC is considered as the fingerprint of a protein and usually the first 2D NMR measurement to start with. Figure 2.6 shows the $^{1}H/^{15}N$-HSQC profile of ubiquitin with full assignments acquired from this study.

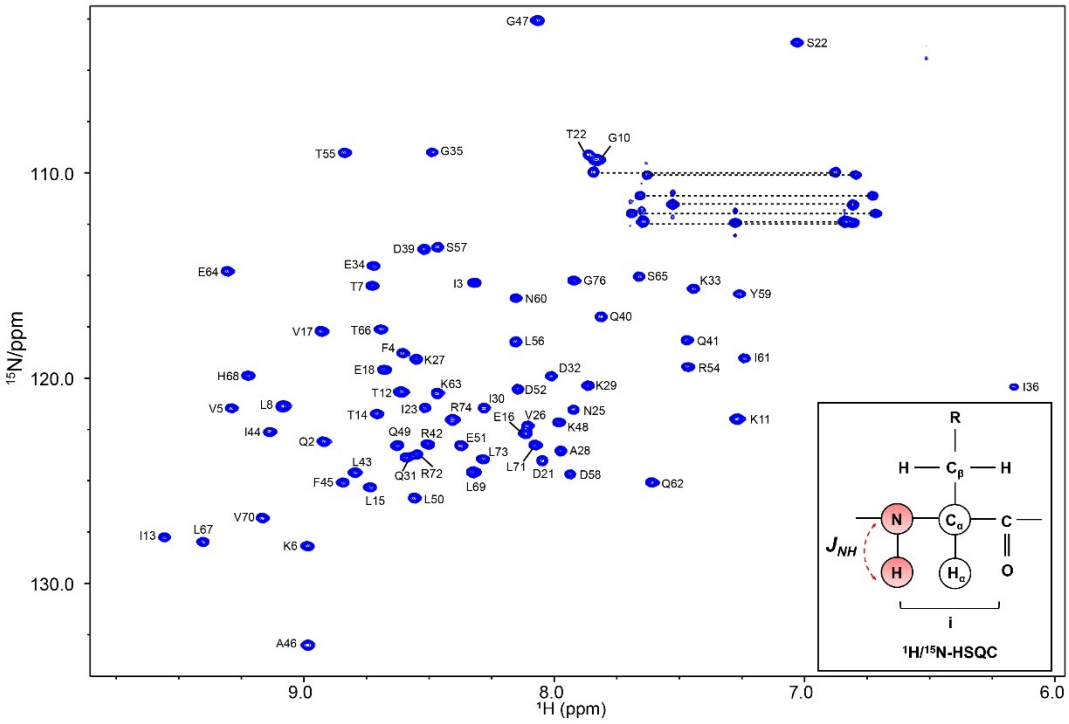


Figure 2.6 The $^{1}H/^{15}N$-HSQC spectrum of ^{15}N-labelled ubiquitin. The assignment for each residue is indicated (expect the side chains). The dashed lines link the signals from the side chains. The insertion in the right corner indicates the J coupling in the NH group. Atom signals visible in the spectrum are coloured in red. $^{1}H/^{15}N$-HSQC spectrum was acquired from this study. The signals were assigned according to the previously reported chemical shifts of ubiquitin (BMRB entry: 17769).

$^{1}H/^{15}N$-HSQC can be combined with three-dimensional ^{15}N-NOESY to obtain the proton-proton constraints for structure calculation. The NH strip of each residue in ^{15}N-NOESY can be identified from $^{1}H/^{15}N$-HSQC. Then any protons that have the NOE effects with the sidechain protons in this NH strip can be detected. Likewise, $^{1}H/^{13}C$-HSQC or triple resonance experiments can also be combined with ^{13}C-NOESY to help obtain the structure constraints.

2.5 Three-dimensional NMR experiments and protein assignment theory

Heteronuclear triple resonances are widely used for protein assignments by solution NMR. The sample is usually uniformly labelled with ^{13}C and ^{15}N, and all the three nuclei are involved in specific NMR detection, resulting in a three-dimensional spectrum. Triple resonances provide rich information to assist in solving the protein structure.

2.5.1 Triple resonance experiments for protein backbone assignments

Taking HNCA, one of the most used triple resonance experiments for protein backbone assignments as an example, the three dimensions of this experiment are ^{1}H, ^{15}N and ^{13}C respectively. The spectrum can be seen as many 2D $^{1}H/^{15}N$-HSQC planes (Figure 2.7B, x-z plane) extending along the third axis ^{13}C (Figure 2.7B, y-axis). Thus, each 2D plane contains the signals at a certain ^{13}C value. When compact all the 2D $^{1}H/^{15}N$-HSQC planes into one plane, it becomes the full view of a normal HSQC spectrum (Figure 2.7C).

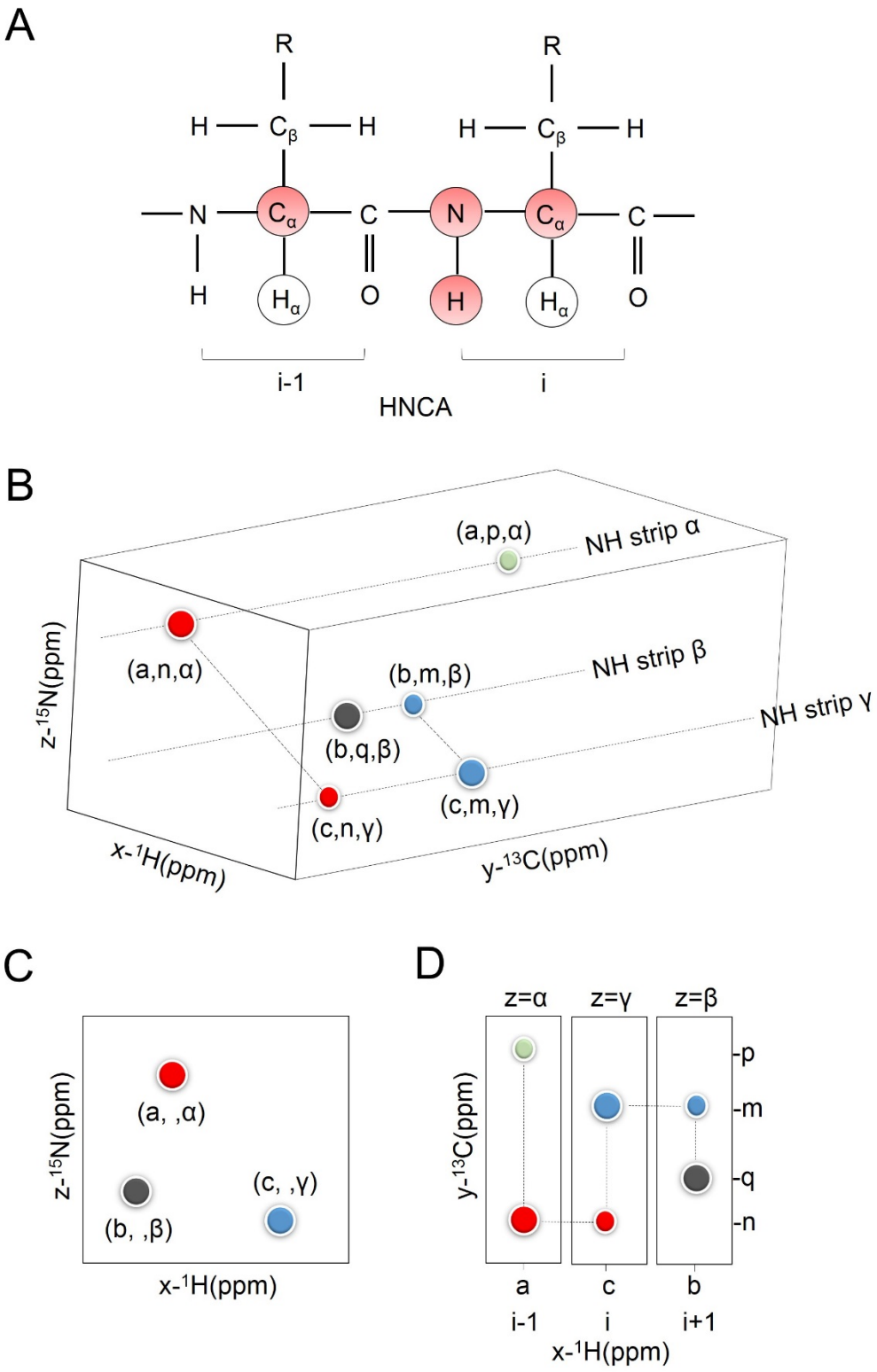


Figure 2.7 Schematic representation of an HNCA experiment. A. The Cα signals of residue i and i-1 are both visible in the HNCA spectrum. Atom signals visible in the spectrum are colored in red. B. A 3-dimensional view of the signals in the HNCA spectrum. The (x,y,z) coordination of each signal is indicated by letters. C. A 2-dimensional view of the HNCA

spectrum by compacting all the x-z planes together. The value of the ^{13}C in y-axis cannot be indicated in this view and thus absent in the coordination. D. Three NH strips at different ^{15}N values (α, β and γ) identify the residue sequence by matching the same signals from different strips.

To analyze the HNCA spectrum in a better way, each 2D $^{1}H/^{13}C$ plane (Figure 2.7B, x-y plane) is usually taken out at a certain ^{15}N value (Figure 2.7B, z-axis) at one time. In each such plane, the signals can be isolated in the form of an "NH strip" (Figure 2.7D). For a certain NH strip of residue i in the HNCA spectrum, the Cα signals of residue i and i-1 are both visible, while the intensity from residue i is stronger than that from i-1 (Figure 2.7A and D). The chemical shift of the Cα signal of a certain residue provides a unique coordination in the 3-dimensional space. By looking for and comparing the NH strips in pairs, the context of two adjacent residues can be identified. Under ideal circumstances, a spectrum like HNCA is enough to identify a series of strips in the correct order of the actual protein sequence. However, multiple factors would possibly affect the resolution of the spectrum in the real experiment. For instance, a bad signal-to-noise ratio would increase the difficulty to distinguish some of the residues. Some signals can be also overlapped. The flexible area of the protein usually displays weak signals.

To help assign the Cα signal from each residue, HN(CO)CA can be used in conjugation with HNCA (Figure 2.8A). For a certain NH strip of residue i in the HN(CO)CA spectrum, only the Cα signals of residue i-1 are visible. If overlay the same NH strip from the two spectra, Cα from residue i-1 can be confirmed (Figure 2.8B). This would help distinguish a real signal from the background noise.

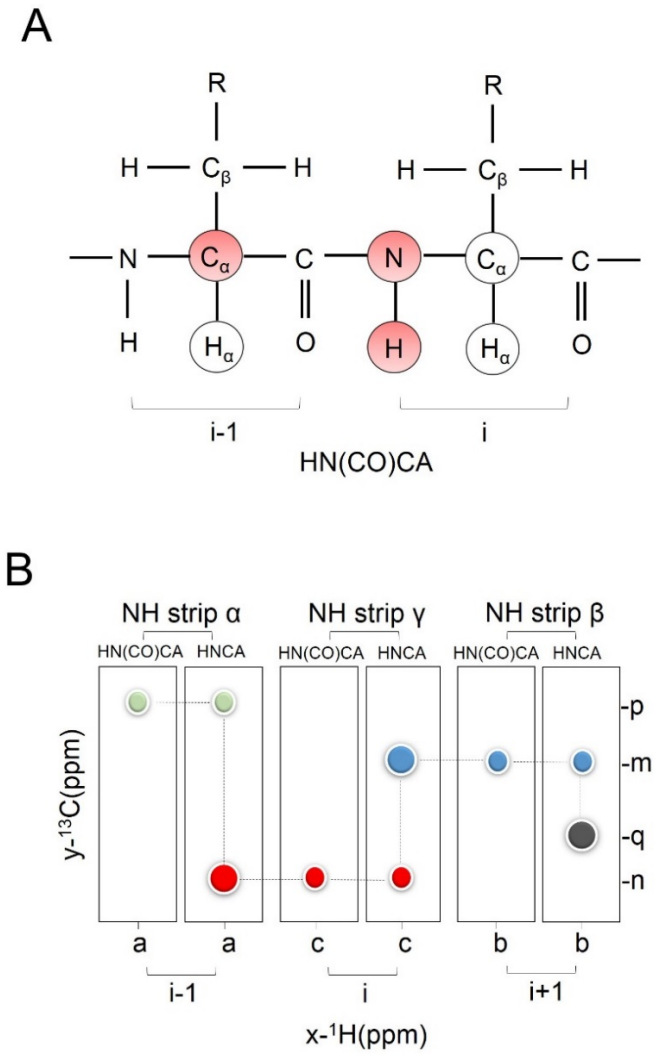
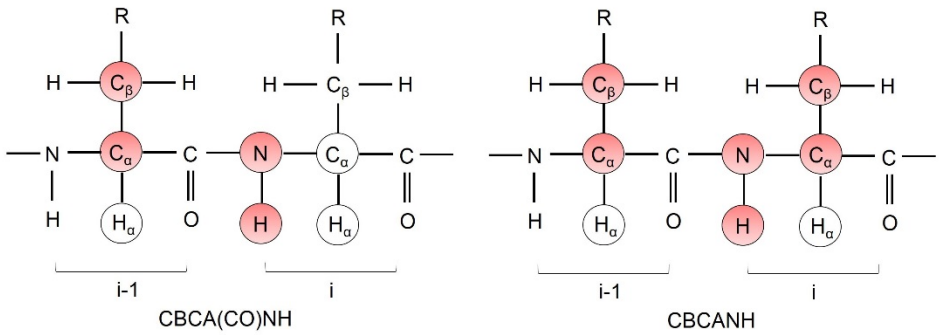


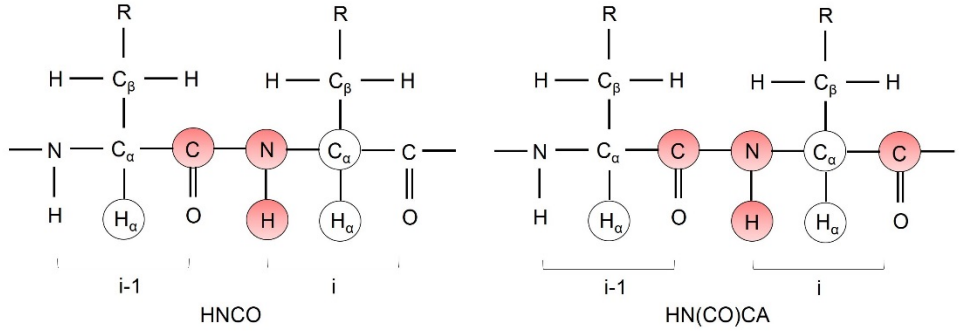
Figure 2.8 Schematic representation of an HN(CO)CA experiment in conjugation with HNCA. A. The Cα signal of residue i-1 is visible in the HN(CO)CA spectrum. Atom signals visible in the spectrum are colored in red. B. By overlapping the same NH strip from HNCA and HN(CO)CA, The Cα signal of residue i-1 can be identified.

There are other spectrum combinations to help with the protein backbone assignments. As shown in Figure 2.9A, CBCA(CO)NH and CBCANH can be conbined together to help identify both Cα and Cβ from residue i-1. It can also work with HNCA/HN(CO)CA to confirm the assignment of Cα. Similarly, carbonyl CO can be identified by combining HNCO and HN(CO)CA at the same time (Figure 2.9B).

A



B



C

Spectrum	Cα		Cβ		CO	
	i-1	i	i-1	i	i-1	i
HNCA	✓	✓				
HN(CO)CA		✓				
CBCA(CO)NH		✓		✓		
CBCANH	✓	✓	✓	✓		
HNCO					✓	
HN(CA)CO					✓	✓

Figure 2.9 Schematic representation of different combinations of triple resonance experiments to assist protein backbone assignments. A. The Cα and Cβ signals can be identified in CBCA(CO)NH and CBCANH experiments. Signals from residue i and i-1 are both visible in CBCANH while only signals from residue i-1 are visible in the CBCA(CO)NH. B. Carbonyl CO can be identified in HNCO and HN(CO)CA experiments. Signals from residue i and i-1 are both visible in HN(CO)CA while only signals from residue i-1 are visible in HNCO. Atom signals visible in the spectrum are colored in red. C. A summary of the visible signals (indicated by the tick marks) from each experiment.

2.5.2 Triple resonance experiments for protein side chain assignments

There are also specifically developed triple resonance experiments to identify the sidechain C and H groups. The most commonly used experiments include HBHA(CO)NH, HCCH-TOCSY etc.

As shown in Figure 2.10A, the Hα and Hβ signals from residue i-1 correlate to the NH group from residue i in HBHA(CO)NH. With the help from the $^{1}H/^{15}N$-HSQC spectrum, the NH strip of each residue can be found in HBHA(CO)NH and the proton signals in this strip are actually from the preceding residue. HBHA(CO)NH is quite sensitive and usually recorded prior to HCCH-TOCSY.

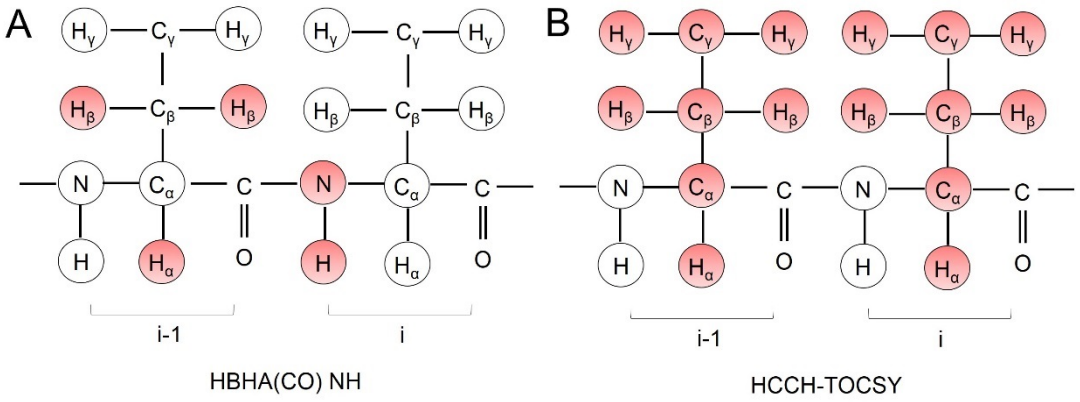


Figure 2.10 Schematic representations of HBHA(CO)NH and HCCH-TOCSY experiments. A. The Hα and Hβ signals from residue i-1 are visible in HBHA(CO)NH. B. Sidechain proton signals from residue i and i-1 are both visible in HCCH-TOCSY. Atom signals visible in the spectrum are colored in red.

HCCH-TOCSY (^{1}H-^{13}C-^{13}C-^{1}H Total Correlation Spectroscopy) is a more complex experiment where the proton resonances from residue i and i-1 can be both detected (Figure 2.10B). These proton signals correlate to the backbone or sidechain carbon signals. For example, in a CH strip of a certain Cα, the Hα signal can be found in the diagonal of the spectrum. In theory, all

the sidechain proton signals from this residue and the preceding residue are visible in this strip. The Hα and Hβ assignments from HBHA(CO)NH and the Cα and Cβ assignments from the previous backbone experiments can be combined with HCCH-TOCSY to help complete almost all the sidechain assignments.

2.6 Protein dynamic studies by NMR titrations

The advantage of using NMR technique to study protein-ligand or protein-protein interactions is that the thermodynamic (Kd), kinetic (k_{on} and k_{off}) and structural (chemical shifts and NOEs) information of the interaction can be well-characterized (Williamson, 2013). In particular, the fast dynamics of a process can be detected by NMR. Another advantage is that a broad range of interaction strengths (from nM to mM) can be detected by NMR, thus it is possible to study weak interactions by this approach. In addition, the characterization is under solution conditions with an atom-level resolution, making NMR an ideal and powerful tool for protein dynamic studies.

A series of $^{1}H/^{15}N$-HSQC experiments are usually employed in the protein-ligand or protein-protein interactions. The protein of interests can be labelled with ^{15}N, in order to be visible in HSQC. Upon the addition of an unlabelled binding partner (i.e. a ligand or a protein) at different ratios, the changes of a variety of physical parameters from the protein can be monitored via NMR through the titration process. The dynamics of the binding partner can be also characterized *vice versa*.

In a simple one-step interaction, in the presence of a binding partner (represented as L here), the line width and Larmor frequency of the signal from a certain residue might be changed compared to the free-bound state (P) (Figure 2.11). These changes are dependent on the exchange rate k_{ex} between the free-bound state and the bound state (PL).

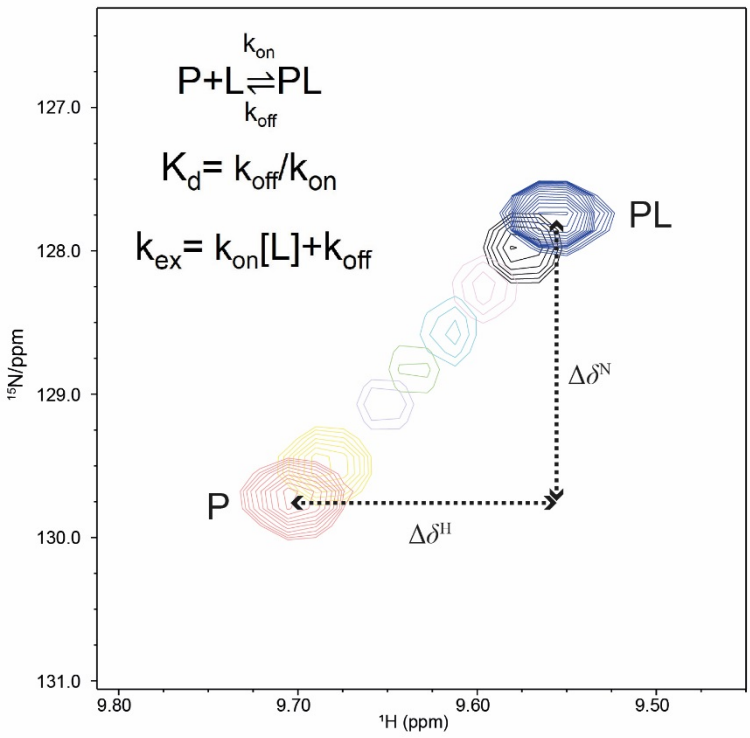


Figure 2.11 Schematic representation of a single step interaction observed from NMR titration. P indicates the free-bound state of the protein of interests. PL indicates the bound state. L represents a binding partner. $\Delta\delta^N$ is the chemical shift difference of ^{15}N. $\Delta\delta^H$ is the chemical shift difference of ^{1}H.

The chemical shift difference $\Delta\delta$ of a signal can be calculated based on the weighted average of $\Delta\delta^H$ and $\Delta\delta^N$. The interaction can be roughly grouped into three regimes depending on the strength of the exchange rate (Figure 2.12). In the case of slow exchange ($k_{ex} << \Delta\delta$), the original signal starts to decrease and then gradually reappears in a new position during the titration process. While for a fast exchange ($k_{ex} >> \Delta\delta$), the signal moves progressively to a new position without much change in the signal intensity. The intermediate exchange is in between the two extreme cases and the line shape and intensity are not changed linearly (Waudby et al., 2016). A series of $\Delta\delta$ collected from the titration can be plotted as a function of L/P ratio, in order to estimate the Kd constant for the interaction.

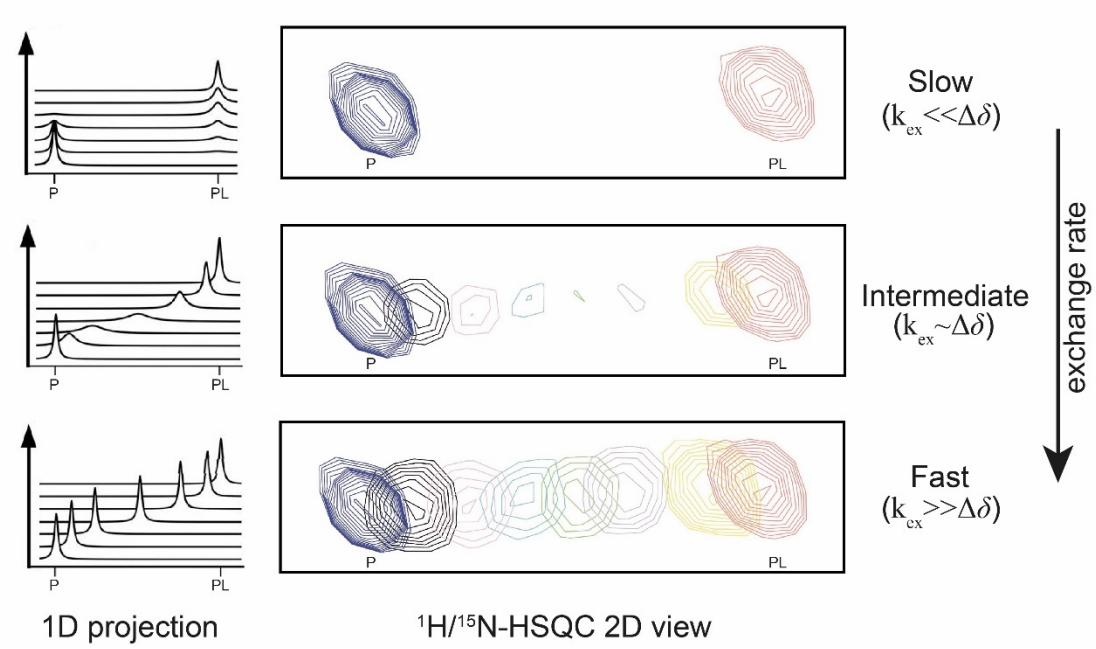


Figure 2.12 Three exchange regimes in the NMR titration process. Left panel: 1D projections of the ^{15}N (or ^{1}H) signals from a residue in ^{1}H/^{15}N-HSQC spectra during titrations. Right panel: the moving patterns of the signal from a residue in ^{1}H/^{15}N-HSQC spectra during titrations. P indicates the free-bound state of the protein of interests. PL indicates the bound state. L represents the binding partner.

If the structure and NMR assignment of the protein (or ligand) are both available, a list of $\Delta\delta$ can be generated summarizing the chemical shift perturbation (CSP) for the binding event. CSPs can be mapped to the structure of the protein or ligand to help identify the binding site or binding interface. The information can be used to establish the computational docking model between the two binding partners.

Apart from the NMR titrations by recording HSQC spectra, NOESY experiments are also sometimes employed to detect intermolecular NOEs that may occur in the binding interface. This could further help identify the binding residues in pairs and support the binding model.

CHAPTER III

Materials and Methods

3.1 Materials

3.1.1 Equipment

General equipment	***Manufacturer/Source***
Balance (ED2202S-CW)	Sartorius
Digital Block Heater	JENCONS-PLS
Mini-PROTEIN® Tetra Electrophoresis System	Biorad
P-1 Peristaltic Pump	Pharmacia Fine Chemials
pH Meter (3510)	JENWAY
Tactrol 2 Autoclave	Priorclave
T100™ Thermal Cycler	Biorad
ThermoMixer C	Eppendorf
Vacuum Filtration Device	Millipore
XCell SureLock Mini-Cell Electrophoresis System	Invitrogen
Centrifuges	
Avanti J-30I	Beckman Coulter
Centrifuge 5810R	Eppendorf
Centrifuge 5424	Eppendorf
Universal 320R	Hettich
Columns	
HiLoad® 16/600 Superdex® 75 pg (SEC)	GE Healthcare
HiLoad® 26/600 Superdex® 75 pg (SEC)	GE Healthcare
HiLoad® 16/600 Superdex® 200 pg (SEC)	GE Healthcare
HiLoad® 16/600 Superdex® 200 pg (SEC)	GE Healthcare
HiTrap® Desalting	GE Healthcare

HiTrap® HP SP (cation exchange)	GE Healthcare
HisTrap™ HP (Ni-affinity)	GE Healthcare
Mono Q 5/50 GL (anion exchange)	GE Healthcare
PD-10 (desalting)	GE Healthcare
Resource S (cation exchange)	GE Healthcare
Self-packed Ni-NTA column (Ni-affinity) (~10 mL, filled with GSH Sepharose™ 4 Fast Flow)	GE Healthcare
Self-packed Ni-NTA column (Ni-affinity) (~10 mL, filled with Ni Sepharose™ 6 Fast Flow)	GE Healthcare
Self-packed TALON column (Co-affinity) (~20 mL, filled with TALON resin)	GE Healthcare
Superose™ 12 10/300L (SEC)	GE Healthcare
Superdex® 200 Increase 10/300 GL (SEC)	GE Healthcare
Imagers	
ChemiDoc™ Touch Imaging System	Biorad
Typhoon TRIOVariable Mode Imager	Biorad
ITC	
MICROCAL PEAQ-ITC	Malvern
Incubators	
Excella E25 shaker	New Bruswick
Innova 44R shaker	New Bruswick
Protein purifiers	
AKTA Prime™	GE Healthcare
AKTA Pure™	GE Healthcare

AKTA Purifier™	GE Healthcare
Spectrometers	
Cary 100 UV-Vis spectrophotometer	Aglient
WPA Biowave II Spectrophotometer	Biochrom
DS-11+ Spectrophotometer	DeNovix
SPR	
Biacore S200	GE Healthcare

3.1.2 Consumables

Amicon® Utra-15 Centrifugal Filters (3K/10K)	Merk Millipore
NuPAGE Novex 4-12% Bis-Tris Gel (12/17-wells)	Invitrogen
Mini Dialysis Kit (1 kDA cut-off, 250 µL)	GE Healthcare
Mini-PROTEAN® TGX Stain-Free™ Precast Gels	Biorad
RunBlue™ 4-12% Bis-Tris Gel (9cm×10cm)	Expedeon
Series S Sensor Chip CM5	GE Healthcare
Shigemi NMR tube	Bruker
SnakeSkin™ Dialysis Tubing	Thermo Scientific
Syringe filter (0.2 µm)	Fisherbrand
VIVASPIN 500 (3,000 MWCO)	Sartorius stedim biotech

3.1.3 Chemicals

Ampicillin sodium	Apollo Sceintific
Bolt™ MES SDS Running Buffer (20×)	Novex
cOmplete™ protease inhibitor cocktail tablets	ROCHE
Cy3 Maleimide Mono-Reactive Dye	GE Healthcare

D-^{13}glucose	CIL
DTT	Apollo Sceintific
gibco®100X MEM vitamin solution	Thermo Fisher Scientific
Instant Blue™	Expedeon
Kanamycin sulphate	Apollo Sceintific
(L)-glutathione (reduced form)	Apollo Sceintific
Lumitein Protein Gel (1×)	Biotium
MES Run Buffer (20×)	Expedeon
Miller's LB Broth	Melford
$^{15}NH_4Cl$	CIL
NuPAGE® LDS Sample Buffer (4×)	Novex
PMSF	Melford
Quick-Load® Purple 1 kb Plus DNA Ladder	NEB
RunBlue™ TriColor Prestained Protein Ladders	Expedeon
TCEP-HCl	Apollo Sceintific

Other general cheimicals are purchased from Apollo Sceintific, Fisher Scientific, Melford, Sigma or VWR.

3.1.4 Kits

Amine Couping Kit	GE Healthcare
GeneJET plasmid Miniprep Kit	Thermo Scientific
GST Capture Kit	GE Healthcare
PCR clean up kit	NEB; QIAGEN
QIAprep Spin Miniprep Kit	QIAGEN
Kit for Molecular Weights (29-700 KDa)	Sigma

3.1.5 Enzymes

KOD Hot Start DNA Polymerase	Novagen
NEBuilder® HiFi DNA Assembly MasterMix	NEB
OneTaq MasterMix	NEB
Q5 Polymerase MasterMix	NEB
Restriction Enzymes (NdeI-HF, EcoRI-HF, BamHI, DpnI)	NEB

3.1.6 Backbone plasmids used in this thesis

Name	Antibiotic resistance
pCold-I	Amp
pCold-TF	Amp
pET47b	Kana
pET49b	Kana
pGEX-6P1	Amp

3.1.7 *E. coli* Competent cells

BL21 (DE3) competent *E. coli*	Self-made
NEB® 5-alpha Competent *E. coli* (High Efficiency)	NEB
Rosetta DE3 competent *E. coli*	Self-made
XL2 competent *E. coli*	Self-made

3.2 Methods

3.2.1 Microbiology-related methods

3.2.1.1 General *E. coli* culture induction

For *E. coli* expression strains carrying pGEX-4T1, pGEX-6P1, pET47b or pET49b-derived vectors, the corresponding LB culture was incubated at 37°C/220 rpm until OD_{600} reached to ~0.8. Then the culture was induced by 0.1-0.5 mM IPTG at 20°C/180 rpm for ~16 h.

To improve protein yield, pCold-I and pCold-TF vectors were used in this study. Both vectors carry the cold-shock Protein A (*cspA*) promoter, allowing the synthesis of the recombinant protein at a low temperature (15°C) in *E.coli*. Under such condition, the expression of host proteins is suppressed and the protease activity is decreased, which helps increase the purity and yields of the recombinant proteins. pCold-TF vector additionally expresses the 48-kDa trigger factor (TF), a prokaryotic ribosome-associated chaperone protein, as a soluble tag fused to the expressed protein of interest. It facilitates co-translational folding of nascent polypeptides, thus reducing misfolding and insolubility of the protein (pCold™ TF DNA user manual, TAKARA).

To this end, for *E. coli* expression strains carrying pCold-I or pCold-TF-derived vectors, the corresponding LB culture was first incubated at 37°C until OD_{600} reached to 0.5-0.8. Then the culture was cold-shocked in the ice/water mixture until the temperature reduced to ~15°C and further kept in 15°C for 30 min without shaking. Later, IPTG was added at the final concentration of 0.1-0.5 mM for induction at 15°C/180 rpm for ~24 h.

3.2.1.2 *E. coli* growth in M9 media for isotopic labelling

Protocol

A small LB culture (~5 mL) was inoculated with a tiny amount of *E. coli* glycerol stock from -80°C to recover the strain overnight. On the second day, the above recovered culture was re-inoculated into a new 5 mL LB medium and incubated for ~ 6 h. Then 200 µL of this small culture was added into 25

mL M9 media as starting culture. The M9 starting culture was incubated overnight at 37°C/220 rpm. The OD_{600} should be 2-3 the next day. Then 25 mL starting culture was added into 1 L M9 media for incubation at 37°C/220 rpm. Note that for a 5L flask, only incubate at most 1L culture to supply enough aeration for good baterial growth. In the case of growing *E.coli* Rosetta DE3 strains to express His-tagged N4BP1 CUE domain, IPTG was added to a final concentration of 1 mM when OD_{600} reached to ~ 0.8 at 37°C and then the induction was conducted at 20°C overnight.

M9 media

To make 1 L M9 media, dissolve 6 g Na_2HPO_4, 3 g KH_2PO_4 and 0.5 g NaCl in 800 mL water and pH to 7.4, then make up to 1 L and transfer into a 5L conical flask for autoclave. Take ~ 40 mL above buffer to dissolve the following reagents, filter through the 0.2 μM membrane and add back to the bulk media:

D-glucose (D-^{13}glucose if ^{13}C labelling is needed)	2 g
NH_4Cl ($^{15}NH_4Cl$ if ^{15}N labelling is needed)	0.7 g
gibco® 100X MEM vitamin solution	10 mL
0.1 M $FeSO_4$	100 μL
1M $CaCl_2$	10 μL
1M $MgSO_4$	2 mL

Add antibiotics to the culture before inoculation.

3.2.1.3 Making *E. coli* competent cells

Protocol

A small 1 mL LB culture with corresponding antibiotic was made from the *E. coli* (XL2 or BL21 or Rosetta DE3 strains) glycerol stock overnight. Then it was used to inoculate a new 10 mL overnight pre-culture. On the next day, 1 mL pre-culture was added to 100 mL LB media and incubated at 37°C/250 rpm until OD_{600} reached to ~0.4 (taking 2-3 hours). The culture was then

transferred into 2×50 mL polypropylene falcon tubes and left on ice for 15 min, then centrifuged at 4°C for 5 min at 5K rpm. The pellet was collected and gently resuspended (avoid vortexing or excessive pipetting) in 25 mL pre-chilled TfbI Buffer and left on ice for 15 min. It was then centrifuged at 4°C for 5 min at 5K rpm. The pellet was collected again and gently resuspended in 4 mL pre-chilled TfbII Buffer. Small aliquots (100 µL each) were made and flash frozen in liquid nitrogen. Aliquots stored at -80°C can be used for transformation for up to 6 months.

Buffers used in this protocol

TfbI Buffer: Dissolve 0.294 g Potassium Acetate, 0.989 g $MnCl_2 \cdot 4H_2O$, 0.745 g KCl, 0.147 g $CaCl_2 \cdot 2H_2O$, 15 mL glycerol in 100 mL Milli-Q H_2O. Filter sterilize with 0.2 µm syringe filter before use.

TfbII Buffer: Dissolve 0.209 g MOPS, 1.1 g $CaCl_2 \cdot 2H_2O$, 0.075 g KCl and 15 mL glycerol in 100 mL Milli-Q H_2O. Filter sterilize with 0.2 µm syringe filter before use.

3.2.2 Molecular biology-related methods

3.2.2.1 Gene cloning

Gene cloning was conducted by using Q5 polymerase (NEB), following manufacturer's instructions. If the PCR product only shows single band with correct size from the TAE agarose gel, the PCR product can be directly purified by PCR Clean-up Kit (NEB). If not, the corresponding band is cut from the gel, then the DNA is purified by gel extraction using the Gel Extraction Kit (Qiagen).

3.2.2.2 Gene ligation

Inserting genes into the corresponding plasmid was all achieved by the In-Fusion cloning technology using NEBuilder® HiFi DNA Assembly MasterMix (NEB).

In brief, the gene specific primer for gene amplification is designed with an extra 15-bp extension that is homologous to the end of the linearized

vector. By PCR, both ends of the gene was added with the corresponding 15 bp vector sequence. With the help from the In-Fusion enzymes, the gene fragment was incorporated into the linearized vector by homologous recombination of the overhangs. A typical reaction is as follows:

Reaction components	Volume
2× NEBuilder® HiFi DNA Assembly MasterMix	5 μL
DNA insertion	~ 200-300 ng
Linearized plasmid	~ 100 ng
H_2O	top up to 50 μL

The reaction was incubated at 50°C for 30 min using a PCR thermocycler. Then 2 μL reaction was taken for the transformation into XL2 or DH5α competent cells. Positive strains were selected under corresponding antibiotic resistance and confirmed by sequencing.

The detailed information of the primers is listed in Appendix IV.

3.2.2.3 Site-directed mutagenesis (SDM)

1) Primer design

The method to achieve the site-directed mutagenesis in this thesis is generally amplifying the whole plasmid by using a pair of primers which carry the mutation. The primer design is based on the online design website (https://www.agilent.com/store/primerDesignProgram.jsp) from Agilent following the instruction step by step.

The detailed information of the primers for SDM is listed in Appendix IV.

2) PCR

KOD Hot Start DNA Polymerase (Novagen) was used to amply the whole plasmid carrying the point mutation. The typical elongation time is ~25 s/kb. The reaction system and condition are shown below. Usually a 2×50 μL reaction is prepared. When the PCR ended, 15 μL reaction was loaded onto a 1%

TAE agarose gel to make sure a visible PCR product is generated.

PCR reaction

Reaction components	Volume
10×KOD buffer	5 μL
25 mM dNTPs	5 μL
25 mM $MgSO_4$	3 μL
10 μM Fwd Primer	2 μL
10 μM Rev Primer	2 μL
Plasmid template (20 ng/μL)	1 μL
KOD polylmerase	1 μL
PCR H_2O	top up to 50 μL

PCR program

Temperature	Time (min)	
95°C	02:00	
95°C	01:00	
50°C	01:00	
70°C	02:30	go to step 2 (25×cycles)
70°C	05:00	
4°C	Pause or forever	

3) DpnI digestion and transformation

The rest of the PCR product (85 μL) was incubated with 3 μL Dpn I (NEB) for digestion at 37°C for 3-4 h or overnight. The DNA was purified by the PCR Clean-up Kit (NEB) with an elution of ~ 20 μL. Then 2 μL DNA was taken for the transformation into the XL2 or DH5α Competent Cells. Positive clones were selected under corresponding antibiotic resistance and confirmed by sequencing.

3.2.3 Protein-related methods

3.2.3.1 General protein purification

A. GST-tagged protein purification and tag removal

Protocol

The *E. coli* cell pellet was defrosted and resuspended with ~ 100 mL Resuspension Buffer, followed by sonication at 75% amplitude with a 60s-on/60s-off pulse on the ice, for 4-5 repeats. The lysate was centrifuged at 20K rpm for 30 min at 4°C. The supernatant was then loaded onto the GSH column (pre-equilibrated with 5 CV Washing Buffer 1). The column was washed with Washing Buffer 1 (high salt) to remove most of the impurities until the UV absorption reached a steady baseline. The column was then washed with Washing Buffer 2 (low salt) to reduce the salt concentration. Finally, the GST-tagged protein with eluted with Elution Buffer containing 20 mM GSH. The protein elution is usually concentrated to 5~10 mL for GST-tag removal.

To remove the GST tag, the protein was transferred into a dialysis bag (SnakeSkin™ Dialysis Tubing, Thermo Scientific) and incubated with the corresponding protease (usually thrombin or HRV-3C protease according to the linker sequence). Then the dialysis bag was dialysed against ~1.5 L Dialysis Buffer for overnight cleavage as well as dialyzing GSH out of the sample. If using thrombin, the cleavage is usually conducted at 22°C or R.T. If using HRV-3C protease, cleavage can be done at 4°C.

After dialysis, the GST column was equilibrated with Washing Buffer 2. The overnight cleavage mixture was loaded through the GSH column again to remove most of the free GST and the un-cleaved GST-tagged protein. In theory, the flowthrough contains mostly the tag-free protein. This flowthrough was again concentrated until the volume was reduced to desired (usually 1-5 mL depending on the loop size for the gel filtration). A gel filtration column with appropriate size was chosen for the size-exclusion

chromatography (SEC) and equilibrated with SEC Buffer. The protein sample was loaded onto the column from the loop and eluted with SEC buffer for the separation. The corresponding peaks from the chromatography were tested by SDS-PAGE and the fractions with pure target protein were pooled to the desired concentration.

Small protein aliquots were made and flash-frozen in the liquid nitrogen and then kept in -80°C.

Buffers used in this protocol

Resuspension Buffer: 500 mM NaCl, 100 mM Tris-HCl (pH 7.4), 1 mM DTT

Washing Buffer 1: 500 mM NaCl, 100 mM Tris-HCl (pH 7.4), 1 mM DTT

Washing Buffer 2: 150 mM NaCl, 100 mM Tris-HCl (pH 7.4), 1 mM DTT

Elution Buffer: dissolve 0.6 g glutathione (GSH) into 100 mL washing buffer 2, check and adjust the pH to 7-8 if necessary

Dialysis Buffer: Washing Buffer 2

SEC Buffer: 150 mM NaCl, 50 mM HEPES (pH 7.6), 1 mM DTT

B. His-tagged protein purification and tag removal

Protocol

The *E. coli* cell pellet was processed using the same procedure as described above. The supernatant after sonication and centrifuging was loaded onto the TALON or Ni-NTA column (pre-equilibrated with 5 CV Equilibration Buffer). The column was then washed with Equilibration Buffer until the UV absorbance reached a steady baseline. To elute the target protein, the column was washed with 0-100% gradient elution from Equilibrium Buffer to Washing Buffer with 5-10 CV. The peak fractions were tested by SDS-PAGE and concentrated to desired. The following tag removal and separation by SEC is the same as what's described in section 3.2.3.1B.

Buffer used in this protocol

Equilibration Buffer: 300 mM NaCl, 50 mM Tris-HCl or HEPES, if necessary, add TCEP-HCl to 0.5 mM. No DTT is allowed if using TALON or Ni-NTA

column.

Washing Buffer: Equilibrium Buffer containing 250 mM imidazole

C. On-Column Cleavage

Apart from the cleavage in the dialysis process, in some cases, the tag can be removed by on-column cleavage.

The *E. coli* lysate was prepared in the same way as stated above. Then the supernatant from the cell lysate after centrifuging was incubated with the pre-equilibrated resin on an orbital rotator for 2-4 h at 4°C. The resin was then collected on an empty column and washed several times with the corresponding buffer to clean off the impurities as much as possible. Later, the resin was resuspended with 1-1.5 CV buffer containing the corresponding protease (usually thrombin or HRV-3C protease) and incubated on an orbital rotator for the cleavage overnight at 4°C. In theory, given enough incubation, the tag would be kept attached to the resin and the free-tagged protein would be released into the buffer. To collect the buffer containing the tag-free protein, the resin was collected on the empty column again and the flowthrough was concentrated for further purification steps.

3.2.3.2 Purification of ubiquitin and variants

There are several different types of ubiquitins used in this study as summarized in Table 3.1, including K48/K63/M1-linked di-ubiquitin chains (K48/K63/M1-diUb), M1-linked tetra-ubiquitin chains (M1-tetraUb) and ubiquitin mutants such as K48A, K48R and M-Cys-Ub (A Cys residue was introduced in between M1 and Q2 of ubiquitin). The purification of tagged ubiquitin proteins follows the standard method as described in section 3.2.3.1.

Here the purification of tag-free ubiquitin and K48/K63-diUb chains will be specified.

Table 3.1 Ubiquitin-related proteins used in this thesis

Name	Tag	Tag removed	Purification method
Ub WT	Tag-free	-	SEC
Ub K48A	GST	Yes	GSH*/SEC
Ub K48R	GST	Yes	GSH*/SEC
M-Cys-Ub	GST	Yes	GSH*/SEC
K48-diUb	Tag-free	-	IEX
K63-diUb	Tag-free	-	IEX/SEC
M1-diUb	His	No	His*/SEC
M1-tetraUb	GST	Yes	GSH*/SEC
NEDD8	GST	Yes	GSH*/SEC

His*: Affinity chromatography by TALON or Ni-NTA column

GSH*: Affinity chromatography by GSH column

SEC: Size-exclusion chromatography

IEX: Ion exchange chromatography by Resourse S column

A. Tag-free ubiquitin purification

Ubiquitin has high tolerance to acidic environment, therefore it can remain soluble at pH 4.0 while most of the other proteins will be precipitated. This is the general method to isolate ubiquitin from a relatively large amount of culture.

For ^{15}N-labelled ubiquitin, the general purification procedure is almost the same, the only difference is to grow the culture in ^{15}N-M9 media in order to label the protein.

1) Purification of ubiquitin by acid-precipitation

The *E. coli* pellet from the LB culture was collected and resuspended in ~ 60 mL SEC buffer (150 mM NaCl, 50 mM HEPES, pH 7.4) and then sonicated at 75% power for 4×1 min. Then the lysate was centrifuged at 4°C/20K rpm for 30 min to collect the supernatant. The pH of the supernatant was adjusted

into 4.0 by adding 1M HCl dropwise in order to precipitate most of the impurities. The protein mixture was centrifuged at 4°C/20K rpm for 30 min to collect the supernatant, most of which contained ubiquitin with only a few impurities. The pH was then adjusted back to 7.4. This neutralized supernatant was concentrated to 2-5 mL using a 3K concentrator (Merk Millipore) in order to load onto the SEC column. It's better to exchange the protein buffer into SEC buffer via the concentrator to decrease the salt strength that was introduced by adjusting the pH.

2) Size-exclusion chromatography (SEC)

Usually 2 mL concentrated neutralized sample was loaded onto the SEC column (HiLoad® 16/600 Superdex® 75 pg, GE Healthcare) and eluted with the SEC buffer. Ubiquitin is only ~ 8.5 KDa, thus eluted at a very late stage. The corresponding peaks from the chromatography were tested by SDS-PAGE and the fractions with pure ubiquitin were pooled to the desired concentration. Protein aliquots were flash frozen in the liquid nitrogen and then kept in -80°C.

B. The synthesis of K48-linked diUb chains

1) Enzymatic formation of K48-diUb chains

The K48-linked ubiquitin chains cannot be synthesized simply by bacterial expression. Instead, an E2 enzyme CDC34 is required to specifically make the K48-linked ubiquitin chains in combination with E1 (Ube1) *in vitro*.

The reaction is prepared in the reaction buffer (50 mM HEPES pH 7.6, 150 mM NaCl, 50 mM $MgCl_2$) as follows. DTT is not allowed in the reaction. ATP addition is the last step to initiate the reaction. The reaction is usually incubated overnight at 25°C. To improve the yield of K48-diUb chains, the reaction time can be optimized to reduce the production of higher Mw chains.

Component	Final concentration
Ube1	1 μM
CDC34	10 μM
Ub	250 μM
ATP	10 mM

2) Purification by IEX (cation exchange)

The 5.0 mL reaction above was mixed with equal amount of 50 mM NaAc (pH 4.0) to precipitate Ube1 and CDC34. Then the mixture was incubated on a 60°C heating block for 30 min followed by centrifuging at 20K rpm for 20 min to remove the protein precipitation. The supernatant, enriched with ubiquitin chains, was then loaded onto the 1 mL Resource S column (GE Healthcare) using the sample pump from the AKTA Pure™ system (GE Healthcare). Before loading the sample, the Resource S column was pre-equilibrated with 50 mM NaAc (pH 4.0). Then the proteins were eluted with 300 CV salt gradient (0-1.0 M NaCl in 50 mM NaAc pH 4.0) with 2mL/fraction. Usually before reaching to 0.5 M NaCl, all the attached protein can be washed out from the column. The K48-diUb is usually the second peak after elution started. The corresponding fractions were pooled and exchanged into the desired buffer required for the downstream experiments.

C. The synthesis of K63-linked diUb chains

1) Enzymatic formation of K63-diUb chains

Similar to the synthesis of K48-linked diUb chains, the K63-diUb chains can be synthesized *in vitro* by two E2 enzyme UBE2V1 and UBC13 working together in combination with E1 (Ube1).

The reaction is prepared in the reaction buffer (50 mM HEPES pH 7.6, 150 mM NaCl, 50 mM $MgCl_2$) as follows. DTT is not allowed in the reaction. ATP addition is the last step to initiate the reaction. The reaction is usually

incubated overnight at 25°C. To improve the yield of K46-diUb chains, the reaction time can be optimized to reduce the production of higher Mw chains.

Component	Final concentration
Ube1	1 μM
UBE2V1	10 μM
UBC13	10 μM
Ubiquitin	250 μM
ATP	10 mM

2) Purification by cation exchange

The purification of K63-diUb chains is almost the same to that of K48-linked diUb chains. Note that the K63-diUb separated by cation exchange is usually contaminated with higher Mw chains. A further SEC (Superdex® 200 Increase 10/300 GL) step was applied to polish the proteins.

3.2.3.3 ^{15}N/^{13}C-labelled tag-free CUE purification

Protocol

The cell pellet harvested from ~ 3 L *E. coli* culture in M9 media was resuspended in TALON buffer and then processed by sonication. The supernatant from the lysate was mixed with the pre-equilibrated ~15 mL TALON resin at 4°C for ~5 h to get better incubation. Then the resin was collected on the column and washed with TALON buffer. The resin was re-suspended and incubated with ~ 2 CV TALON buffer containing 150 μL 27 mg/mL HRV-3C protease for the on-column cleavage overnight at 4°C.

The second day, the resin was collected on the column and the flowthrough was enriched mostly with the cleaved protein. The flowthrough was pooled to ~ 2 mL and then loaded onto the SEC (HiLoad® 16/600 Superdex® 75 pg, GE Healthcare). The proteins were eluted with NMR buffer. The cleaved CUE protein (~5.2 KDa) has very low UV absorption and eluted

at a very late stage. The peak fractions were confirmed by SDS-PAGE and the pure fractions were pooled to ~ 300 μL at 500-800 μM including 10% D_2O and 0.05% NaN_3 for NMR assignments as well as the NMR titrations.

Buffer used in this protocol

TALON buffer: 300 mM NaCl, 50 mM Tris-HCl, 0.5 mM TCEP-HCl

NMR buffer: 22 mM phosphate, 55 mM NaCl, pH 7.0, 1mM DTT

3.2.3.4 Purification of chimeric RNAse and RNAseCUE

Protocol

The procedures to purify chimeric RNAse or RNAseCUE were optimized into 3 steps. The protein was first purified by TALON affinity chromatography as described in section 3.2.3.1 except that the protein was directly eluted with Washing Buffer containing 250 mM imidazole. The peak fractions were combined and pooled to 2.5 mL for desalting using the pre-equilibrated PD10 column (GE Healthcare) by IEX Buffer A. The 3.5 mL elution was then loaded to the pre-equilibrated Mono Q column (1 mL, GE Healthcare). The column was washed with 0-100% gradient elution from IEX Buffer A to IEX Buffer B with a total 100 CV. The pure chimeric protein was eluted at a salt concentration of ~100 mM NaCl. The peak fractions from IEX were combined and pooled to ~100-200 μL for analytical size-exclusion chromatography using Superose™ 12 10/300L (GE Healthcare). Protein was eluted with GF buffer.

Buffer used in this protocol

IEX buffer A: 20 mM HEPES (pH 8.5), 20 mM NaCl, 1 mM DTT

IEX buffer B: 20 mM HEPES (pH 8.5), 1 M NaCl, 1 mM DTT

GF Buffer: 150 mM NaCl, 50 mM HEPES (pH 8.5), 1 mM DTT

Note: All the buffers above contain 0.5% Triton X-100.

3.2.3.5 Protein quantification

Protein concentration was measured on a DS-11+ Spectrophotometer (DeNovix) using microlitre volume under A_{280} mode. The absorption at 280

nm was applied to the Beer-Lambert's law for the concentration calculation.

$$A = \varepsilon \cdot c \cdot l \quad (1)$$

Where:

A is the absorbance value

ε is the molar extinction coefficient of the measured protein (M^{-1} cm^{-1})

c is the protein concentration (M)

l is the path length (1 cm)

The AKTA protein purifier also monitors the purification progress at 280 nm by the UV detector based on the above principle. Note that for the analytical size-exclusion chromatography in section 3.2.5.1, Triton X-100 was used in the buffer. It is known that Triton X-100 also has strong absorption at 280 nm. To minimize the interference from Triton X-100, a background subtraction was performed before analysing the protein sample. The SEC column was equilibrated with the buffer until the UV absorbance reached steady. The UV absorbance was then manually adjusted to 0 by the AutoZero function from the UNICORN software.

3.2.3.6 SDS-PAGE

Protein samples were mixed with equal amount of NuPAGE® LDS Sample Buffer (2×, containing 0.5 M DTT) and denatured at 95°C for 10 min. Usually a 10 μL sample was loaded to the RunBlue™ 4-12% Bis-Tris Gel (Expedeon). The protein gel was run in 1×MES Running Buffer (Expedeon) at 140V for ~ 1.5 h. Then the gel was stained with Instant Blue™ (Expedeon) for ~ 10 min to visualize the protein bands. The gel image was taken by the ChemiDoc™ Touch Imaging System (Biorad).

3.2.3.7 Proteins purified in this thesis

Table 3.2 summarizes the basic information and purification methods of the proteins used in this thesis. The tag of some of the proteins were kept for certain applications, such as SPR immobilization or to help increase the protein solubility. His tag is usually un-cleaved unless specified.

Table 3.2 Proteins purified in this thesis

Protein name	Mw (KDa)	Tag removed	Purification method
CDC34	27.6	No	His*/SEC
Chimeric N4BP1 RNAse domain	23.0	No	His*/SEC
Chimeric N4BP1 RNAse + CUE domain	36.1	No	His*/SEC
HOIL-1L (1-478)	54.5	Yes	GSH*/SEC
HOIP (697-1072)	45.3	No	His*/SEC
HOIP (300-1072)	86.9	Yes	GSH*/SEC
HRV-3C protease	19.9	No	GSH*/SEC
N4BP1 (1-311)	34.8	Yes	GSH*/SEC
N4BP1 (1-342)	37.9	Yes	GSH*/SEC
N4BP1 (1-392)	43.9	Yes	GSH*/SEC
N4BP1 CUE domain (850-893)	53.9	No	GSH*/SEC
(including WT and F862G/P893A)	5.2	Yes	His*/SEC
N4BP1 KH-like domain (1-144)	17.9	No	His*/IEX
N4BP1 RNAse domain (613-774)	71.1	No	His*/SEC
N4BP1 RNAse + CUE domain (613-893)	84.2	No	His*/SEC
N4BP1 UBA-like domain (312-392)	11.5	No	His*/SEC
N4BP1 UBM-like domain (144-311)	21.2	No	His*/SEC
UBC13	20.3	No	His*/SEC
UbcH5c	16.7	Yes	GSH*/SEC
Ube1	117.8	No	His*/SEC
UBE2V1	22.6	No	His*/SEC

His*: Affinity chromatography by TALON or Ni-NTA column

GSH*: Affinity chromatography by GSH column

SEC: Size-exclusion chromatography

IEX: Ion exchange chromatography by Resourse S column

3.2.4 Biochemical methods

3.2.4.1 Generation of Cy5-labelled ubiquitin

Ubiquitin is labelled with Cy5 maleimide dye (GE Healthcare) for fluorescence detection in the ubiquitination assay. Cy5 maleimide dye usually reacts with the cysteine thiol group for the labelling. A Cys residue was introduced in between M1 and Q2 of ubiquitin (M-Cys-Ub) to enable labelling.

M-Cys-Ub was dialyzed against the Labelling Buffer (150 mM NaCl, 50 mM HEPES pH 7.6) overnight before the reaction. A 5 to 10-fold excess of Cy5 Maleimide (dissolved in DMSO) was added into 2 mM M-Cys-Ub in the Labelling Buffer containing 0.5 mM TCEP-HCl. The reaction was conducted in the dark at R.T. for 2 h. To terminate the reaction, 20 mM glutathione was added to inactivate the excess dye. The free dye was then removed via a 5 ml HiTrap desalting column (GE Healthcare). The Cy5-labelled ubiquitin was further purified by SEC (HiLoad® 16/600 Superdex® 75 pg, GE Healthcare).

3.2.4.2 Ubiquitination assay

A typical *in vitro* ubiquitination assay usually contains E1, E2 and E3. Tag-free mono-ubiquitin is commonly used as the substrate. A tiny amount of Cy5-labelled ubiquitin (~ 1 μM) is added to the reaction in order to participate in forming the chains, thus, the chains can also be labelled with Cy5. The reaction was prepared in the assay buffer (50 mM HEPES pH 7.6, 150 mM NaCl, 50 mM $MgCl_2$). Note that DTT is not allowed in the reaction. ATP was the last component added to initiate the reaction. The reaction temperature is 25°C.

If using the longer HOIP protein construct (including the UBA domain) for the assay, HOIL-1L is needed to release the auto-inhibition of HOIP. Generally, the final concentration of HOIL-1L in the reaction is 5 μM. To test the inhibition effect from N4BP1, the corresponding N4BP1 construct is added at a final concentration of 25 or 50 μM.

Component	Final concentration
E1 (Ube1)	1 μM
E2 (UbcH5c)	10 μM
E3(HOIP)	1 μM
WT Ubiquitin	100 μM
Cy5-Ubiquitin	1 μM
ATP	10 mM

The samples were collected at different time points during the reaction process and mixed with equal amount of NuPAGE® LDS Sample Buffer (2×, containing 0.5 M DTT) and denatured at 95°C for 10 min. Usually 10 μL sample was loaded to the NuPAGE Novex 4-12% Bis-Tris Gel (Invitrogen). The protein gel was run at 140V for ~ 1.5 h. After the eletrophoresis, the gel was directly scanned at an excitation wavelength of 633 nm and an emission wavelength of 670 nm in a Typhoon TRIOVariable Mode Imager (Biorad) to image the Cy5-labelled ubiquitin. To image the total protein bands, the gel was further stained with Instant Blue™ (Expedeon) for ~ 10 min. The gel images were analyzed by the Image Quant TL software (GE Healthcare).

3.2.5 Biophysical methods

3.2.5.1 Protein molecular weight determination

1) Analytical size-exclusion chromatography

In order to characterize the oligomeric states of the protein, analytical size-exclusion chromatography was performed. The SEC column used in this thesis is Superose™ 12 10/300L (GE Healthcare). The void volume (Vo) of this column is 7.2 mL. Six standard proteins with known Mw were loaded respectively to the pre-equilibrated column and eluted with the GF Buffer (150 mM NaCl, 50 mM HEPES (pH 8.5), 1 mM DTT, 0.5% Triton X-100). Note that Triton X-100 was added to the buffer in order to reduce protein precipitation in the case of chimeric RNAse and RNAseCUE. The

Triton X-100 concentration has been optimized to 0.5%. Due to that Triton X-100 was used in the buffer, a background subtraction was performed as follows before loading the protein onto the column. The SEC column was thoroughly equilibrated with the GF buffer until the UV absorbance reached steady. The UV absorbance was then manually adjusted to 0 by the AutoZero function from the UNICORN software. It is known that 0.5% Triton X-100 exists as micelles at ~90 KDa (CMC=~0.02%). But there was no particular peak of Triton X-100 in any profile of the standard proteins (for example Carbonic Anhydrase or Thyroglobulin), indicating that the column was well-equilibrated and the background had been successfully subtracted.

The standard proteins were taken from the Gel Filtration Markers Kit (29,000-700,000 Da, Sigma). The loading volume for each protein was all at 200 μL. The total loading amount for each standard protein is as follows:

Protein	Mw (Da)	Loading amount
Carbonic Anhydrase, bovine erythrocytes	29,000	1.5 mg
Albumin, bovine serum	66,000	5 mg
Alcohol Dehydrogenase, yeast	150,000	5 mg
β-Amylase, sweet potato	200,000	5 mg
Apoferritin, horse spleen	443,000	5 mg
Thyroglobulin, bovine	669,000	5 mg

The elution volume (Ve) of each protein peak (with maximum UV absorption) from the gel filtration profile was recorded. The calibration curve was plotted using the elution volume/void volume ratio (Ve/Vo) versus the protein molecular weight (in the logarithmic form). It was then linearly fitted. The protein of interest can be analyzed in the same way to obtain the Ve/Vo ratio from the gel filtration profile. Then the molecular weight of the protein can be calculated by applying the Ve/Vo ratio to the calibration curve.

2) SEC-MALLS

SEC-MALLS (Size-exclusion chromatography coupled multiple-angle laser light scattering) is the technique to determine the absolute molecular mass of a protein separated by SEC.

Protein was prepared at 0.5, 1.0, 2.0 mg/mL and then 100 uL of each concentration was separately applied to the pe-equilibrated SEC column (Superdex® 200 Increase 10/300 GL, GE Healthcare) using the buffer containing 150 mM NaCl, 50 mM HEPES pH 7.6, 0.5 Mm TCEP-HCl and 3 mM NaN_3. The scattered light intensity of the elution from SEC was recorded at 16 different angles by the Dawn HELEOS-II and rEX Optilab systems and the concentration was measured using the rEX Optilab differential refractometer (Wyatt Technology). The average molecular mass was calculated by the ASTRA software (Wyatt Technology). Experiments and data analysis were performed with the kind help from Dr. Simone Kunzelmann (The Francis Crick Institute).

3.2.5.2 Isothermal Titration Calorimetry (ITC)

A physical interaction or chemical reaction is always accompanied with the release or absorption of heat. ITC (Isothermal Titration Calorimetry) is the ideal technique with high sensitivity to study the above thermodynamics especially for biomolecules. During titration, ITC directly monitors the enthalpy changes (ΔH°), binding affinity (Ka) and binding stoichiometry (N) in the binding process. Thus, the Gibb's free energy (ΔG°) and the entropy changes (ΔS°) can be calculated according to the below equations. Apart from the above thermodynamic parameters, ITC could also provide kinetic parameters (K_m, K_{cat} etc). This powerful technique has been widely used in quantitatively characterizing interactions among protein, nucleotide, lipid etc and studying enzymatic kinetics.

The relation between the several parameters is as follows:

$$\Delta G^{\circ} = -R \cdot T \cdot \ln Ka \quad (2)$$

$$\Delta G^{\circ} = \Delta H^{\circ} - T \cdot \Delta S^{\circ} \quad (3)$$

Where:

ΔG° is the Gibb's free energy (kcal/mol)

R is the universal gas constant (1.987 $cal \cdot K^{-1} \cdot mol^{-1}$)

T is the absolute temperature in Kelvin

Ka is the association constant (M^{-1})

The dissociation constant Kd is converted from Ka as shown below:

$$Kd = 1/Ka \quad (4)$$

Modern ITC calorimeters usually employ the power compensation method. In the resting state, the sample cell and the reference cell are kept in identical temperature thus no power difference is detected. When the titration starts, injectant is titrated from the syringe into the sample cell which contains the titrand. The enthalpy in the sample cell is changed over the titration process. The heat controller of the sample cell then quickly adjusts the temperature to maintain constant with the reference cell. This power difference is thus recorded (Figure 3.1).

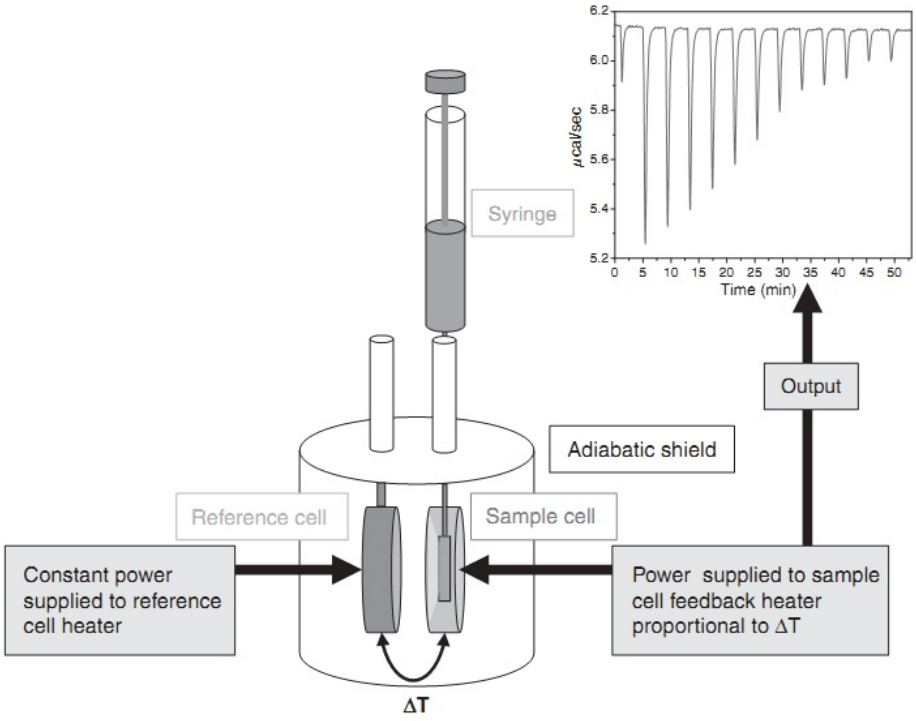


Figure 3.1 The simplified structure of an ITC instrument (© 2008 Freyer and Lewis, by permission, Licence No: 4900970610059). Three major components are shown: the reference cell, the sample cell and the syringe.

The raw data directly acquired from the instrument records the power (µCal/s) versus time during the titration (Figure 3.2A). It is then processed in the form of an isotherm curve where each injection peak is integrated and plotted against molar ratio between injectant and titrand (Figure 3.2B). Thus, several important binding parameters can be calculated when an appropriate fitting model is applied to the isotherm.

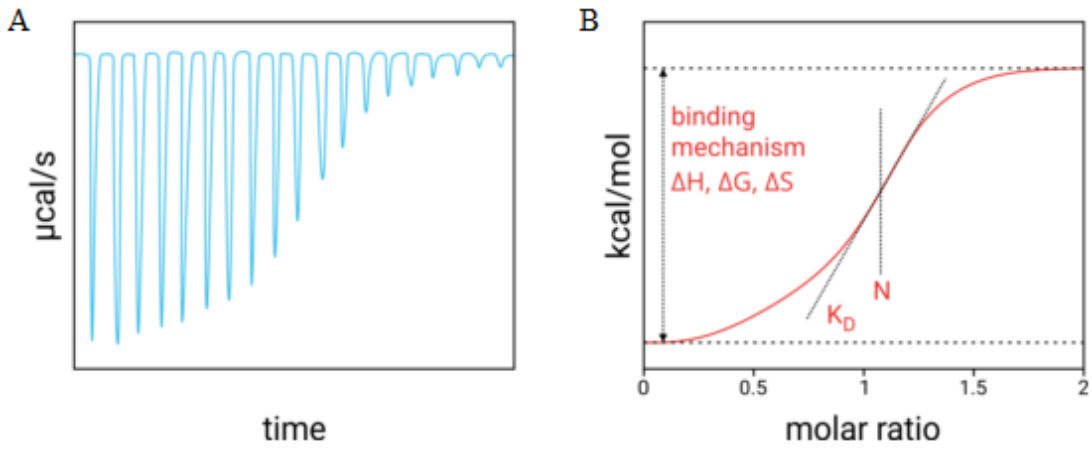


Figure 3.2 An example of an exothermic reaction from an ITC measurement showing the raw data (A) and the isotherm (B). Figure taken from https://2bind.com/itc/.

To fit the isotherm curve, a binding model must be assumed and then an iterative nonlinear regression procedure is used. Due to that all the ITC data in this thesis were using the one-site binding model by the MicroCal PEAQ-ITC Analysis Software, here the below data fitting procedure is described only for this model (Freyer and Lewis, 2008; MicroCal PEAQ-ITC analysis software user manual (MAN0576-01-EN-00)). Equation 5 and 6 describe the relationships between the equilibrium and mass balance in the system. Substituting $[L]$ from Equation 6 into Equation 5 allows the calculation of the fraction of binding site j occupied by the ligand.

$$\Theta_j = \frac{[L]K_j}{1+[L]K_j} \tag{5}$$

$$L_t = [L] + P_t \sum_{j=1}^{k} (n_j \Theta_j) \tag{6}$$

Where:

Θ_j is the fraction of binding site j occupied by the ligand

K_j is the binding constant of process j

L_t is the total ligand concentration

$[L]$ is the free ligand concentration

P_t is the total macromolecule concentration

n_j is the total stoichiometric ratio for process j

Equation 7 describes the total heat content Q of the system. Derived from Equation 7, The differential heat ΔQ_i (the heat for the injection, i represents the injection number) can be thus derived shown as Equation 8.

$$Q=P_tV_o\sum_{j=1}^{k}(n_j\Theta_j\Delta H_j) \quad (7)$$

$$\Delta Q_i=Q_i-Q_{i-1} \quad (8)$$

Where:

Q is the total heat content

V_o is the initial volume of the sample cell

ΔH_j is the molar enthalpy change for process j

Parameter n_j, K_j and ΔH_j are initially estimated by the software and then the nonlinear regression is performed to fit for ΔQi from each injection and compared with the experimental data. The error function is calculated to improve the values of n_j, K_j and ΔH_j for the next round of iteration. The procedure is repeated until no significant improvement is achieve.

Protocol

All the protein samples were dialysed against ITC buffer (50 mM HEPES pH 7.4, 50 mM NaCl, 0.5 mM TCEP-HCl) overnight at 4°C using the Mini Dialysis Kit (1 KDa cut-off, GE Healthcare) before measurements. The titrand in the cell is usually prepared at a range of 100-500 μM. The injectant concentration in the syringe is usually 10-15 fold greater than the titrand. The measurement temperature is 25°C unless specified, with a reference power at 10 μCal/s. Stir speed is 750 rpm. Initial delay is 60s. Standard 19-injection titration program is as follows:

1st injection: 0.4 μl/injection, duration: 0.8s, spacing time: 150 s.

The rest of the injections: 2 μl/injection, duration: 2s, spacing time: 120 s.

Depending on the titrand or injectant types as well as the concentration, the parameters from the titration programs can be adjusted or modified to get better signals. All the ITC measurements were conducted in a MicroCal

PEAQ-ITC calorimeter (Malvern). The data is processed and analyzed using the MicroCal PEAQ-ITC Analysis Software with a one-site binding model (Version: 1.21).

3.2.5.3 Surface Plasmon Resonance (SPR)

Biacore™ systems (GE Healthcare) take the advantage of surface plasmon resonance (SPR) to quantitatively measure the interaction between a wide range of molecules from small chemicals to biomolecules such as nucleic acids, proteins etc. This label-free technique provides kinetics and affinity parameters and has been widely used in drug and antibody screening.

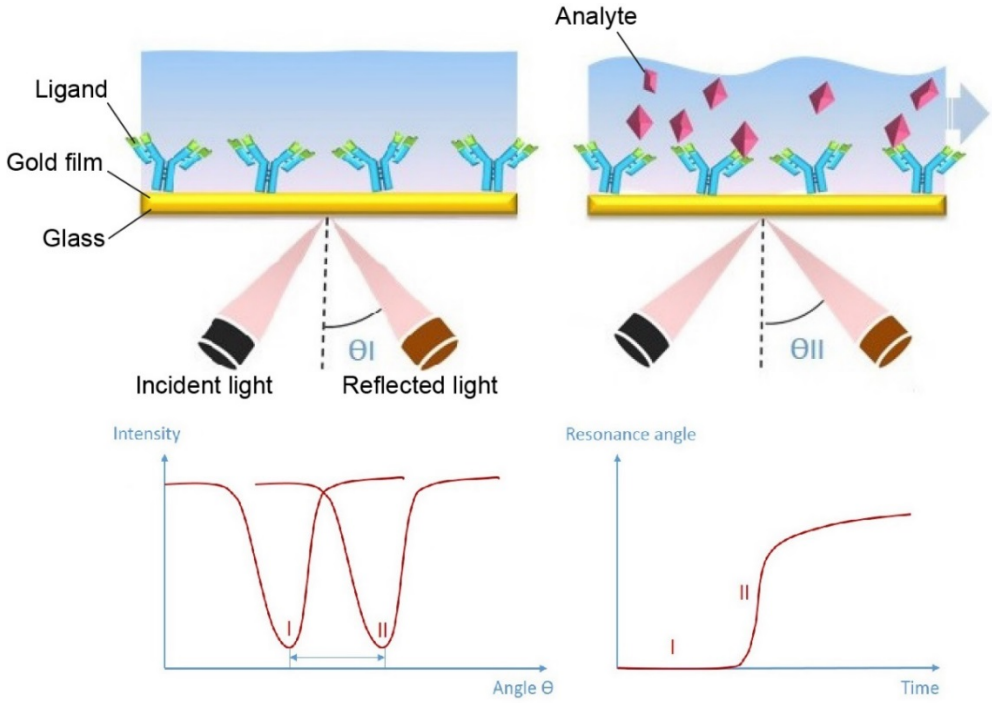


Figure 3.3 The basic SPR detection principle. A SPR sensor chip is a small piece of glass coated with a layer of thin gold film. Ligand is immobilized on the gold sensor surface. Analyte is in the mobile phase and running through the ligand surface under measurement mode. Θ indicates the resonance angle (SPR angle) at different surface conditions. Figure slightly modified from http://weistron.com.cn/.

Surface plasmon resonance is a physical phenomenon that occurs in the interface between two media with different refractive index. Here, the sensor

chip for the Biacore™ system is used as an example. The chip is a small piece of glass coated with a thin film of gold as shown in Figure 3.3. When the incident light is passing from the optically denser medium (i.e. the glass) to the less dense medium (i.e. the gold film), if the incident angle reaches to (or beyond) the critical angle, total internal reflection occurs on the interface. Under a certain incident angle and wavelength, the free electrons (the "surface plasmon") on the gold surface are excited by the incident light and absorb the energy from the incident light, thus the surface plasmon resonance occurs. Because of the energy absorption, the reflected light intensity is greatly reduced. The incident angle under such condition is called the resonance angle or SPR angle. When mass changes on the sensor surface, the local refractive index of the gold layer is changed and results in changes of the SPR angle.

The molecule immobilized on the sensor surface is usually defined as the ligand while its binding partner in the mobile phase is defined as the analyte. Molecules such as proteins can be immobilized on the surface via covalent chemical reactions, or via a capture molecule (such as antibody) to non-covalently interact with the protein with high affinity.

The directly recorded SPR signal over time is displayed in a sensorgram (Figure 3.4). The signal intensity is expressed as response and measured in resonance units (RU). It is directly proportional to the biomolecule mass concentration on the chip surface.

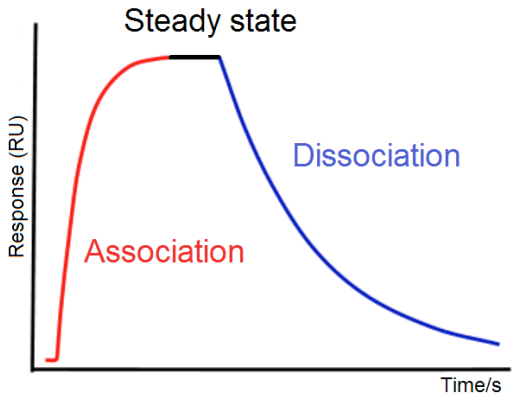


Figure 3.4 A simplified SPR sensorgram.

For an actual measurement, the analyte at a certain concentration is applied over the ligand surface on the sensor chip at constant flow. The association event between the ligand and the analyte increases the mass of the surface, thus results in the increase of the response. For the dissociation event, *vice versa*.

The K_{on} and K_{off} can be resolved from the association and dissociation stage in the sensogram. However, in some cases, the kinetics are too fast to measure, instead the Kd can be resolved by the steady-state method. Here we only described affinity determination at the steady state.

As shown in the sensogram, the steady state phase indicates the equilibrium of analyte association and dissociation with the ligand at a certain analyte concentration. When the ligand surface is fully saturated by the analytes, the response reaches to maximum (Rmax). The response at the equilibrium state under serial analyte concentrations follows the below equation:

$$Req=Ka\cdot c\cdot Rmax/(Ka\cdot c+1) \quad (9)$$

Where:

Ka is the equilibrium (steady state) association constant

Req is the response intensity in the equilibrium (steady state)

c is the corresponding concentration of the analyte (monoUb or diUb chains)

The equilibrium (steady state) dissociation constant K_D is converted from Ka as shown below:

$$K_D=1/Ka \quad (10)$$

<u>Protocol</u>

All the SPR measurements were performed on a Biacore™ S200 (GE Healthcare).

1) GST-tagged protein immobilization

To determine the K_D between GST-tagged CUE domain (hereafter as GST-CUE) with ubiquitin and chains, GST-CUE was immobilized onto the CM5 sensor chip through the polyclonal anti-GST antibody. The CM5 chip was first immobilized with the antibody by amine coupling using a GST capture kit and an amine coupling kit (both from GE Healthcare) following manufacturer's instructions. The polyclonal anti-GST antibody carries a minor fraction of high affinity sites that are difficult to regenerate. To avoid capture of GST-CUE on these sites, prior to the immobilization with GST-CUE, the high affinity site of the antibody was blocked with recombinant GST.

GST-CUE was diluted to 5 μg/mL with 1×HBS-T+ buffer (150 mM NaCl, 10 mM HEPES pH 7.4, 0.05% Tween 20) and loaded to the sample channel in the CM5 chip at a flow rate of 10 μL/min for 1 min. A reference channel was also prepared without loading the protein as control.

2) TF-tagged protein immobilization

To determine the K_D between TF-tagged RNAse or RNAseCUE proteins with ubiquitin and chains, the proteins were directly immobilized onto the CM5 sensor chip by amine coupling using an amine coupling kit (GE Healthcare). For amine coupling, the protein needs to be positively charged in the buffer below its pI for successful immobilization. In our case, the TF-tagged protein stocks (50 μg/mL) were diluted to 5 μg/mL with 10 mM Sodium Acetate at pH4.0, 4.5, 5.0, 5.5 respectively for the pH scouting. pH4.5 was found to perform the best immobilization. Then the TF-tagged proteins were immobilized onto the corresponding sample channels at pH4.5 at a flow rate of 10 μL/min for 20 s.

3) Measurements and data analysis

The analytes (ubiquitin and chains) were dialyzed against 1×HBP-T+ buffer respectively before measurements. A serial dilution of the analyte was

prepared on a 96-well plate. Then the analytes were flowed through the chip at a flow rate of 40 μL/min with 20 s association and 20 s dissociation. The signal from the reference channel was subtracted from the signal obtained from the sample channel.

The corresponding data points were collected at the steady state stage as demonstrated in Figure 3.4 and analyzed by employing the steady-state method described above. The data were fitted using Equation 9 by Matlab.

Experiments and data analysis were performed with the kind help from Dr. Simone Kunzelmann (The Francis Crick Institute).

3.2.5.4 NMR measurements

The basic NMR theory is introduced in Chapter II.

The protein samples are all kept in the NMR buffer (pH 7.0) containing 10% D_2O and 0.05% NaN_3. Protein concentration is 300-500 μM. 300 μL sample was filled in the Shigemi NMR tube (Bruker) for the NMR measurements.

1) NMR measurements for protein assignments

The 2D and 3D NMR spectra of the $^{15}N/^{13}C$-labelled tag-free CUE domain were measured in a Bruker Avance III HD 700 spectrometer equipped with a 5mm $^{1}H/^{13}C/^{15}N$ triple-resonance PFG cryoprobe in the Medical Research Council (MRC) Biomedical NMR Centre (The Francis Crick Institute).

^{15}N-HSQC, HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO, HBHA(CO)NH and HCCH-TOCSY spectra were recorded respectively at 303.15 K under the kind instructions from Dr. Alain Oregioni.

2) NMR titrations

The titrations between the ^{15}N-labelled CUE domain with non-labelled di-ubiquitin chains was measured in a Bruker AvanceNEO 600 spectrometer equipped with a 5mm $^{1}H/^{13}C/^{15}N$ inverse triple resonance probe in the high field Nuclear Magnetic Resonance laboratory of QMUL. A series of ^{15}N-HSQC spectra were recorded under the kind instructions from Dr. Harold Toms.

The titrations between the ^{15}N-labelled CUE domain with non-labelled mono-ubiquitin and the titrations between ^{15}N-labelled di-ubiquitin chains with the non-labelled CUE domain were measured in a Bruker Avance III HD 700 spectrometer equipped with a 5mm ^{1}H/^{13}C/^{15}N triple-resonance PFG cryoprobe in the Medical Research Council (MRC) Biomedical NMR Centre (The Francis Crick Institute). A series of ^{15}N-HSQC spectra were recorded under the kind instructions from Dr. Alain Oregioni.

The CUE domain titrated with ubiquitin

For the titration between the CUE domain and mono-ubiquitin, a series of 0-4.0X titration pairs are made by addition of mono-ubiquitin (stock at 13.17 mM) to the ^{15}N-labelled CUE domain (stock at 0.39 mM).

For the titration between the CUE domain and M1-diUb, a series of 0-5.0X titration pairs are made by addition of M1-diUb (stock at 4.05 mM) to the ^{15}N-labelled CUE domain (stock at 0.4 mM).

For the titration between the CUE domain and K48-diUb, a series of 0-2.0X titration pairs are made by addition of K48-diUb (stock at 2.72 mM) to the ^{15}N-labelled CUE domain (stock at 0.4 mM).

For the titration between the CUE domain and K63-diUb, a series of 0-1.35X titration pairs are made by addition of K48-diUb (stock at 1.465 mM) to ^{15}N-labelled CUE domain (stock at 0.4 mM).

Ubiquitin titrated with the CUE domain

For the titration between mono-ubiquitin and the CUE domain, a series of 0-4.0X titration pairs are made by addition of the CUE domain (stock at 5.27 mM) to ^{15}N-labelled mono-ubiquitin (stock at 0.43 mM).

For the titration between M1-diUb and the CUE domain, a series of 0-3.0X titration pairs are made by addition of the CUE domain (stock at 1.66 mM) to ^{15}N-labelled M1-diUb (stock at 0.3 mM).

For the titration between K48-diUb and the CUE domain, a series of 0-5.0X titration pairs are made by addition of the CUE domain (stock at 1.66 mM) to ^{15}N-labelled K48-diUb (stock at 0.22 mM).

For the titration between K63-diUb and the CUE domain, a series of 0-3.0X titration pairs are made by addition of the CUE domain (stock at 1.66 mM) to ^{15}N-labelled K63-diUb (stock at 0.3 mM).

3) NMR data analysis

The 1D NMR raw data were processed by the Topspin software (Bruker). All the 2D and 3D NMR raw data were first processed by NMRPipe software for data conversion and phase correction. Then the processed spectra were analyzed by CCPNMR software for backbone and side-chain assignments. A total of 97% of the potential backbone (disregarding the proline residues) and 87% of the potential side-chain resonances were assigned (the first 3 N-terminal residues from the tag are ignored).

The HSQC titration spectra were analyzed by CCPNMR software and visualized by NMRviewJ software. Based on the protein assignments, the chemical shift values (δ^N and δ^H) were extracted from the spectra and integrated to the chemical shift perturbation $\Delta\delta$ according to the below equation in order to make the CSP list and map to the structure surface.

$$\Delta\delta = ((\delta^H{}_{bound}-\delta^H{}_{free})^2 + ((\delta^N{}_{bound}-\delta^N{}_{free})/a)^2)^{1/2} \quad (11)$$

$$a = (\delta^N{}_{max}- \delta^N{}_{min})/ (\delta^H{}_{max}- \delta^H{}_{min}) \quad (12)$$

The CSPs were mapped to the protein structure surface by the Chimera software (Version 1.14). The Kd of each residue was calculated by fitting the chemical shift data via the MestReNova software using the advanced tool AFFINImeter-NMR. It also provides an average Kd for the whole titration process. AFFINImeter-NMR applies the stoichiometric equilibrium (SE) models for the NMR titration analysis. For a simple interaction between molecule A and B as shown below, the equilibrium constant can be described as Equation 13:

$$a{\cdot}A+b{\cdot}B \leftrightarrow A_aB_b$$

$$K_{ab}= \frac{[A_aB_b]}{[A]^a[B]^b} \quad (13)$$

For a particular observable X, such as the chemical shift perturbation $\Delta\delta$, it is related to the formation of A_aB_b complex in a volume V and follows the below equation where X_{mn} is the molar contribution of A_mB_n to the observable X:

$$X= \sum_{m,n=0}^{a,b} X_{mn}[A_mB_n]V \quad (14)$$

The formation of A_mB_n can be calculated by Equation 13 where K_{mn} is the global stoichiometric equilibrium constant corresponding to the formation of A_mB_n:

$$[A_mB_n]=K_{mn}[A]^a[B]^b \quad (15)$$

The total concentration of A and B is distributed between the existing free fractions and the complex formed as follows:

$$[A]_T= \sum_{m,n=0}^{a,b} m[A_mB_n] \quad (16)$$

$$[B]_T= \sum_{m,n=0}^{a,b} n[A_mB_n] \quad (17)$$

Where:

$[A]_T$ and $[B]_T$ are the total concentration of A and B

m and n is the numbers of molecule A and B in the A_aB_b complex

X_{mn} and K_{mn} can be obtained by fitting using the above equations to the experimental data via AFFINImeter-NMR (Piñeiro et al., 2019).

The above analysis was performed with the kind instructions from Dr. James Garnetts (King's College London) and Dr. Alain Oregioni (The Francis Crick Institute).

4) Solution NMR structure calculation

The backbone chemical shift data (Cα, Cβ, C', N, Hα and HN) from the complete chemical shift list (Appendix I) was saved in a TALOS format and then submitted to the CS-ROSETTA server (Rosetta version 3.8; https://csrosetta.bmrb.wisc.edu/csrosetta/submit) using the default

parameters. A total of 20,000 models were generated. The top 10 models with the lowest energy were chosen as the final ensemble. Structural statistics (Appendix V) were calculated using several servers including wwwPDB, MolProbity and PROSESS. Favorable Ramachandran statistics were observed, with 100% of residues in most favored (98%) regions and 0% in outlier regions. The whole calculation procedure was performed with the kind instructions from Dr. James Garnetts (King's College London).

5) Establishment of the docking model

The N4BP1^{CUE}/ubiquitin model was built with HADDOCK 2.4 using the default parameters (https://wenmr.science.uu.nl/haddock2.4). The active residues for the ambiguous interaction restraints (AIRs) were defined on the basis of the CSP lists generated from NMR titrations and the relative solvent accessible surface area (SASA) of the residues. The relative SASA was calculated using GetArea (http://curie.utmb.edu/getarea.html). The CSPs and SASA values were combined for the final selection with an in-house R script. The structure for the mono-ubiquitin used here is PDB: 1UBQ. The structure of N4BP1^{CUE} is the top 1 model generated from CS-ROSETTA. The whole modelling process was performed with the kind help from Dr. Arianna Fornili (QMUL).

3.2.6 Bioinformatic methods

3.2.6.1 Protein sequence analysis

Pairwise protein sequence alignment was performed by EMBOSS Neddle (https://www.ebi.ac.uk/Tools/psa/emboss_needle/). Multiple protein sequence alignment was performed by ClustalW (https://www.genome.jp/tools-bin/clustalw) or MEGA (Version 10.0.5) using the default MUSCLE method. Results were visualized by Bioedit.

3.2.6.2 Protein structure analysis

The protein structure prediction was performed by submitting the protein query sequence to the protein fold recognition server Phyre2

(http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) using the intensive modelling mode or by the *de novo* protein structure prediction server trRosetta (https://yanglab.nankai.edu.cn/trRosetta/).

Phyre2 is highly dependent on the known protein structures in order to build the predicted models. The query protein sequence is applied to the multiple sequence alignments (MSAs) to predict secondary structure. It is further combined as a query hidden Markov model (HMM) for the scanning against the HMM database of known structures. The top alignments will be used to establish the initial backbone models where the insertion and deletions will be further corrected by loop modelling. For the final step, the side chains are added to build the final model (Kelly et al., 2015). Apparently, if homology is poor or undetectable between the query and the sequence of a known structure, the modelling will be very unreliable or even fail.

As a newly-developed protein prediction method, trRosetta makes use of MSAs to predict relative distances and orientations of not only residue pairs but also inter-residues of the protein based on a deep residual-convolutional network. Then the predicted distances and orientation information is used to build the *de novo* protein structure with more accuracy following the constrained minimization protocol via Rosetta (Yang et al., 2020). Rosetta performs Monte Carlo search through space of conformations to find minimal energy conformation instead of using homologous structures as templates (Rohl et al., 2004).

Protein structure comparison was performed by submitting the query protein structure to the Dali server for the comparison against the Protein Data Bank (PDB) (http://ekhidna2.biocenter.helsinki.fi/dali/).

The pairwise structure comparison was performed by Pymol using the align function.

CHAPTER IV

Results and Discussion

Chapter 4.1

Qualitative analysis of HOIP inhibition by N4BP1

4.1.1 Introduction: The N-terminus of N4BP1

As mentioned in section 1.4.1, N4BP1 can be cleaved by CASP8 into N- and C-terminal fragments. The N-terminus but not the C-terminus still retains the ability to negatively regulate the TNFR signaling pathway. N4BP1 could also interact with LUBAC core catalytic subunit HOIP. Since LUBAC is a key component in the TNFR signaling pathway as well as the only known E3 ligase that catalyzes linear ubiquitin chain formation, a possible inhibitory role of N4BP1 on the LUBAC E3 ligase activity was proposed and will be investigated in this chapter.

The N-terminus of N4BP1 can be divided into three domains by bioinformatic analysis (Figure 4.1.1A). The KH-like domain is predicted to be involved in RNA-binding, in line with the reported function of N4BP1 in RNA degradation. The UBM-like and UBA-like domain are both putative UBDs, which implicate the possible involvement of N4BP1 in ubiquitin-related processes.

So far, only the structure of KH-like domain has been solved (PDB: 6Q3V). However, the literature is still in the "to be published" state. As can be seen from Figure 4.1.1B, interestingly, the whole crystal structure consists of $N4BP1_{1\text{-}188}$ which almost contains both the KH-like domain ($N4BP1_{6\text{-}144}$) and UBM-like domain ($N4BP1_{160\text{-}200}$) defined here. From the structure, the KH-like domain has a "di-KH" organization, where two subdomains (KH_1 ($N4BP1_{1\text{-}67}$) and KH_2 ($N4BP1_{70\text{-}142}$)) share highly similar fold (RMSD=0.401) though with poor sequence similarity (Figure 4.1.1C).

The interaction between N4BP1 and the LUBAC subunit HOIP has been previously identified by pull-down assays (Figure 4.1.1D, E) (Kliza et al., unpublished data). The identified binding regions of the two proteins are shown in Figure 4.1.2.

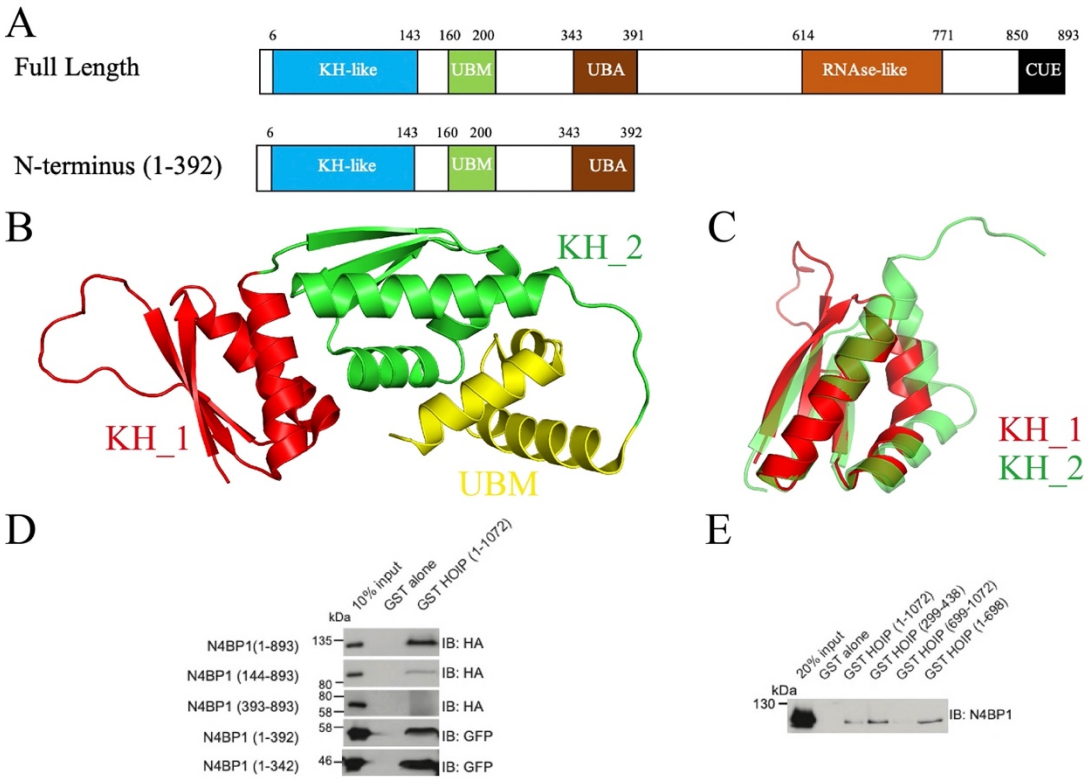


Figure 4.1.1 Current background knowledge of N4BP1 N-terminus. A. The domain structure of the N-terminus and full length N4BP1. Please see Figure 1.28 for N4BP1 domain annotation. B. Crystal structure of N4BP1 KH-like domain (PDB: 6Q3V). The two subdomains are indicated as KH_1 ($N4BP1_{1-67}$) and KH_2 ($N4BP1_{70-142}$). UBM domain ($N4BP1_{143-188}$) is also indicated. C. Superimposition of KH_1 and KH_2. D. $N4BP1_{133-342}$ is identified as the minimum HOIP-binding region. E. $HOIP_{299-438}$ is identified as the minimum N4BP1-binding region.

4.1.2 Experimental design

In order to get mechanistic insights into the inhibitory role of N4BP1 in the ligase activity of LUBAC, *in vitro* ubiquitination assay was performed (Figure 4.12A). Different N-terminal N4BP1 constructs were examined in a series of HOIP-mediated linear chain synthesis (Figure 4.1.2B). N4BP1 was also incubated with HOIP constructs of various length (Figure 4.1.2C). The comparison of linear ubiquitin chain synthesis activity derived from different HOIP constructs in combination with various N4BP1 domains will help expand our knowledge on the inhibitory mechanisms.

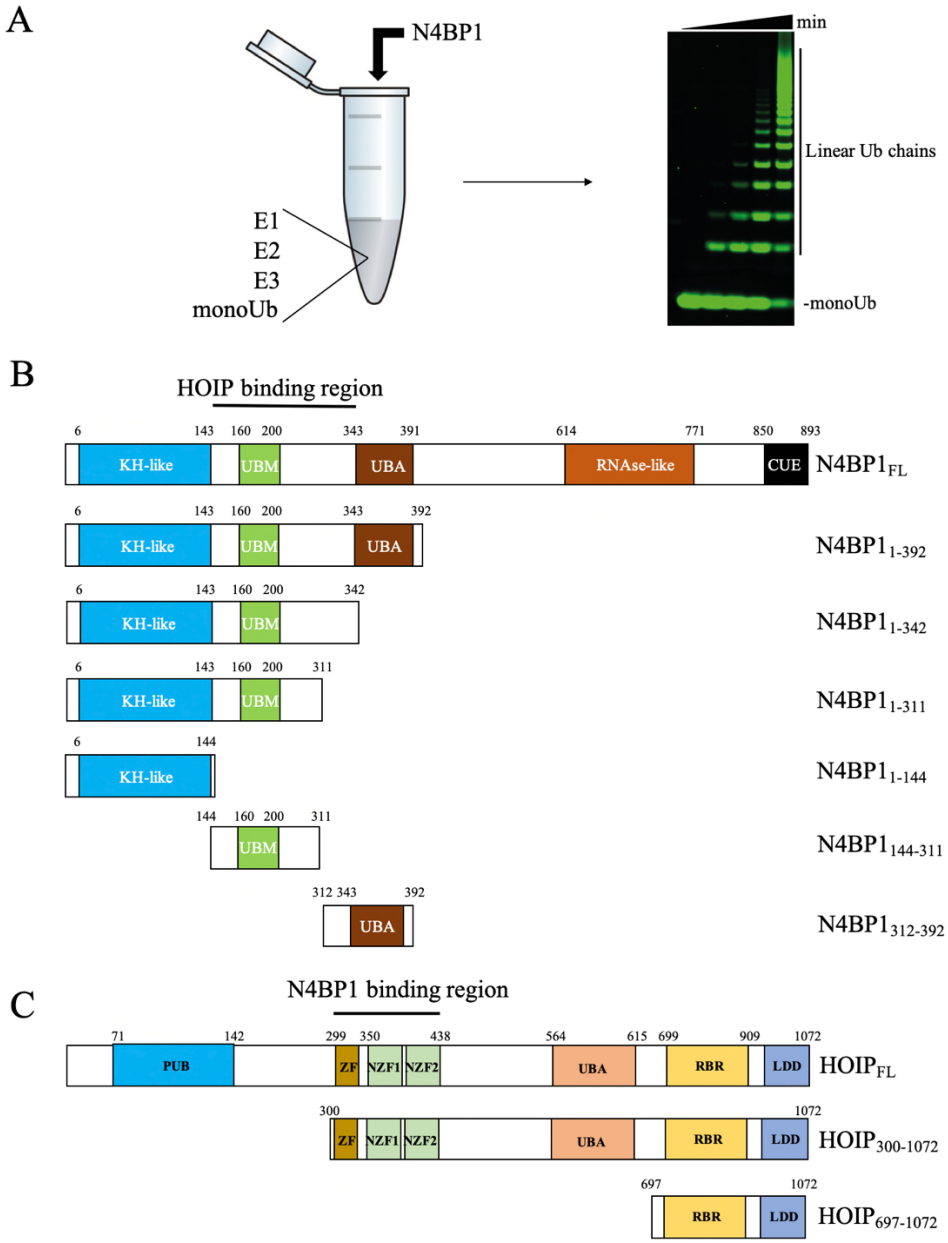


Figure 4.1.2 Experimental design of Chapter 4.1. A. Schematic representative of the *in vitro* ubiquitination assay. N4BP1 was added into the HOIP-mediated linear ubiquitin synthesis system containing monoUb, E1 (Ube1), E2 (UbcH5c), E3 (HOIP). monoUb was labelled with Cy5 in order to visualize the newly-synthesized linear ubiquitin chains. B. The N4BP1 constructs used in this thesis. C. The HOIP constructs used in this thesis. Binding regions between N4BP1 and HOIP are indicated in the full-length protein respectively. LDD: linear ubiquitin chain determining domain. NZF: novel zinc finger. PUB: PNGase/UBA or UBX-containing proteins. RBR: RING-in-between-RING. UBA: ubiquitin-associated. ZF: zinc finger. Please see Figure 1.28 for N4BP1 domain annotation.

4.1.3 Results

4.1.3.1 The N4BP1 KH-like domain displays a weak inhibitory effect on the E3 ligase activity of HOIP catalytic domain

An *in vitro* ubiquitination reaction system containing the minimal key components including ATP, mono-ubiquitin, Ube1 (E1), UbcH5c (E2), HOIP (E3) and N4BP1 (inhibitor) was performed. A HOIP construct comprising residues 697-1072 was used which does not contain the N-terminal domains including the UBA domain which auto-inhibits the ligase activity of HOIP. There are only the catalytic RBR domain and the C-terminal LDD domain present in this construct (Figure 4.1.2A, Figure 4.1.3 upper panel). 1 µM Cy5-labelled mono-ubiquitin was added to an excess of unlabeled ubiquitin which allows to monitor linear ubiquitin chain formation by fluorescent scanning of SDS-PAGE samples taken at different time points.

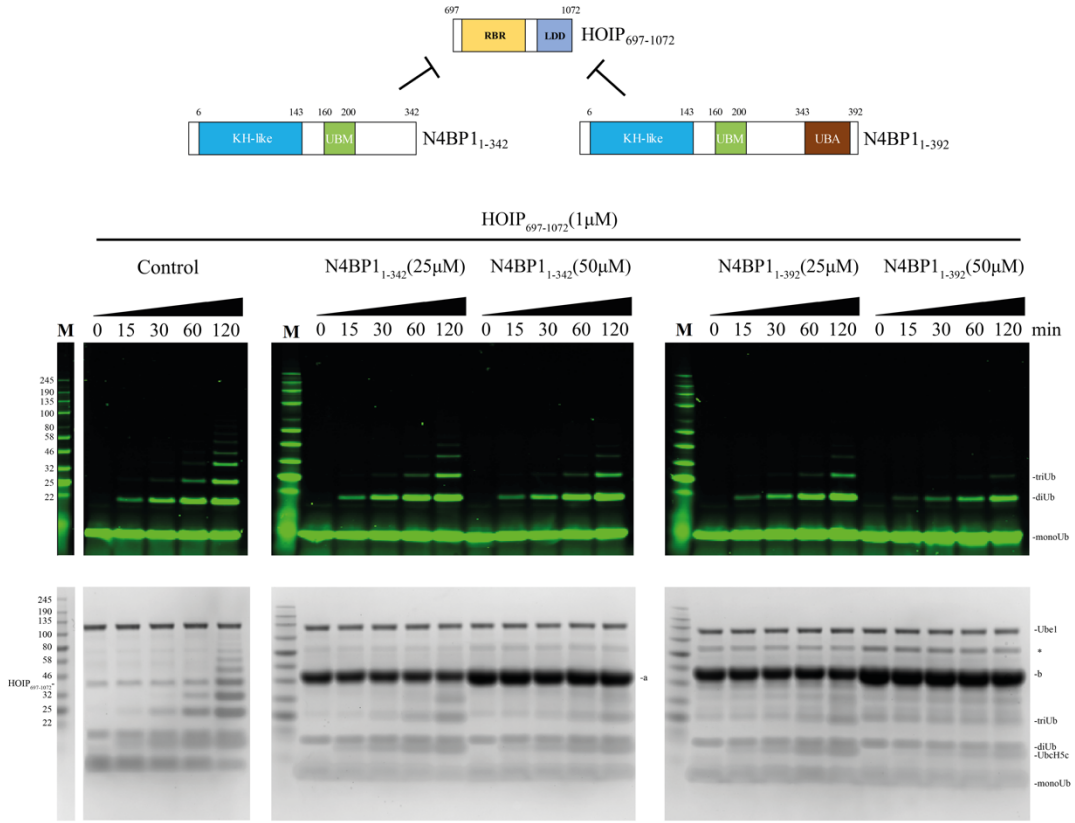


Figure 4.1.3 The inhibitory effect of N4BP1$_{1-342}$ and N4BP1$_{1-392}$ on HOIP$_{697-1072}$. Upper panel: Schematic presentation of the combination of HOIP and N4BP1 constructs in each reaction. Middle panel: The *in vitro* ubiquitination assay results by SDS-PAGE. The gel was scanned at an excitation wavelength of 633 nm and an emission wavelength of 670 nm in order to visualize Cy5-labelled ubiquitin. Bottom panel: Corresponding SDS-PAGE gel stained with Instant Blue™ to visualize the total protein content. The *in vitro* ubiquitination reaction contains: 1 μM Ube1 (E1), 10 μM UbcH5c (E2), 1 μM HOIP$_{697-1072}$, 25 μM or 50 μM N4BP1 construct (the protein band of N4BP1$_{1-342}$ is indicated as "a" while N4BP1$_{1-392}$ is indicated as "b"), 50 μM mono-ubiquitin, 1 μM Cy5-labelled mono-ubiquitin. As a control, there was no N4BP1 added into the reaction. Assay buffer for the reaction is described in detail in section 3.2.4.2. Reaction was initiated by adding ATP at a final concentration of 10 mM. Reaction temperature: 25 °C. M: standard protein ladder. * indicates the impurity band from purifying the N4BP1 constructs.

In the *in vitro* ubiquitination assay, construct N4BP1$_{1-392}$ comprising all the three N-terminal domains shows weak inhibition when compared to the intrinsic activity of HOIP$_{697-1072}$. A similar inhibitory effect is observed in context of a shorter N4BP1 construct (N4BP1$_{1-342}$) which lacks the UBA-like domain (Figure 4.1.3). From this result, it is inferred that this weak inhibitory effect is mediated by a region of N4BP1 which contains the KH-like domain and the UBM-like domain.

To further confirm this observation, the E3 ligase activity of HOIP$_{697-1072}$ was separately monitored in the presence of all three individual N-terminal domains from N4BP1 (i.e. KH-like domain (N4BP1$_{1-144}$) or UBM-like domain (N4BP1$_{144-311}$) or UBA-like domain (N4BP1$_{312-392}$)) (Figure 4.1.4). These results further confirm that the N4BP1 UBA-like domain does not exhibit an inhibitory effect on HOIP$_{697-1072}$. Furthermore, the separation of N4BP1$_{1-342}$ into KH-like and UBM-like domains demonstrates that the KH-like domain mediates the inhibitor function (Figure 4.1.4B). When the reaction is sampled

over shorter period which covers chain synthesis over 60 minutes, it becomes apparent that the KH-like domain has subtle inhibitory effect on linear ubiquitin chain formation catalyzed by $HOIP_{697-1072}$ (Figure 4.1.4C). The UBM-like domain is not involved in the inhibition (Figure 4.1.4B).

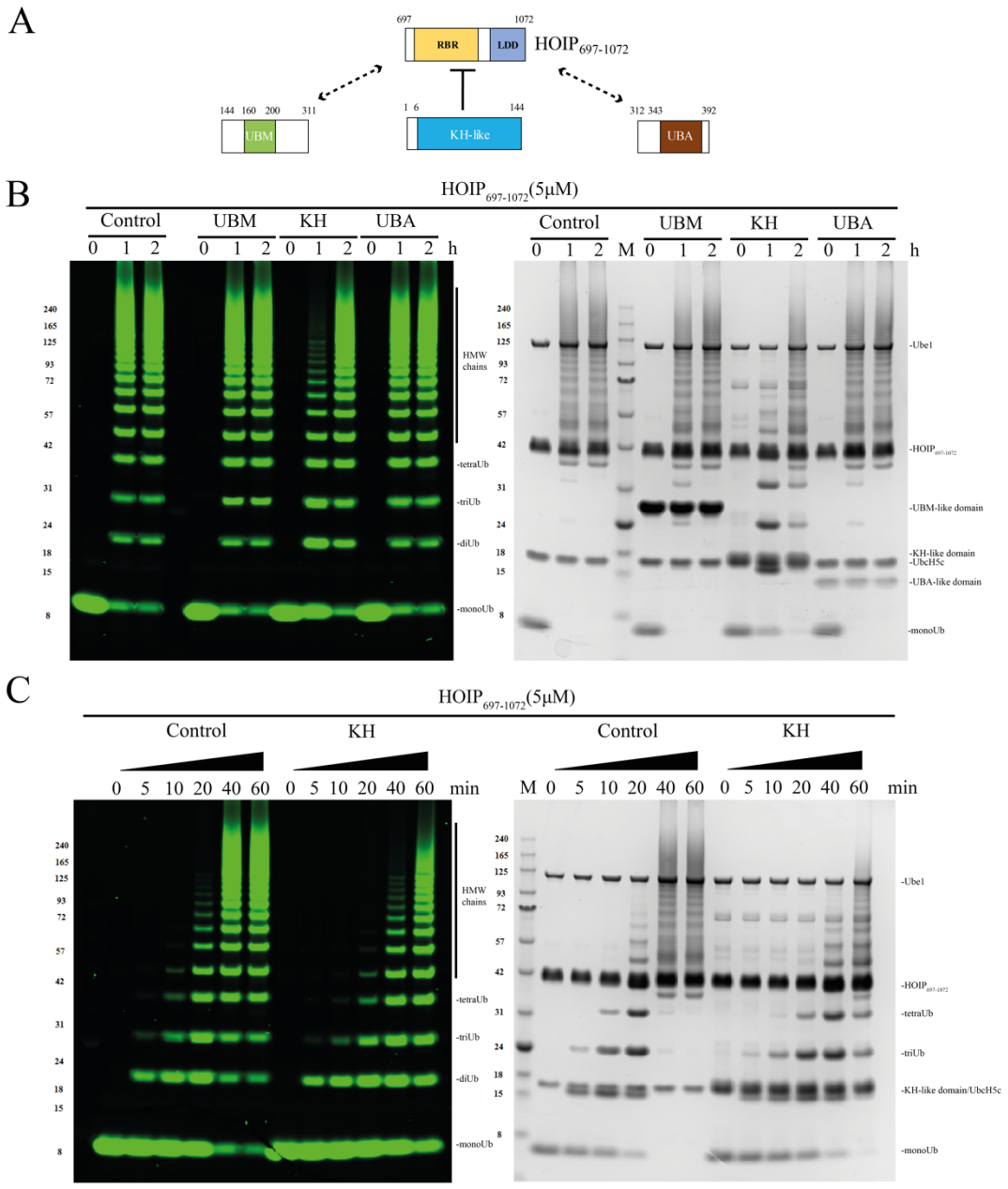


Figure 4.1.4 The inhibitory effect of N4BP1 KH-like domain, UBM-like domain and UBA-like domain on $HOIP_{697-1072}$. A. Schematic of the combination of HOIP and N4BP1 constructs in each reaction. B. The results of *in vitro* ubiquitination assay of N4BP1 KH-like domain, UBM-like domain and UBA-like domain on $HOIP_{697-1072}$ at the indicated 3 time points. C. The results of *in vitro* ubiquitination assay of N4BP1 KH-like domain on

$HOIP_{697-1072}$ at the indicated time points from 0-60 min. Left panel in B and C: The *in vitro* ubiquitination assay results by SDS-PAGE scanned at an excitation wavelength of 633 nm and an emission wavelength of 670 nm in order to visualize Cy5-labelled ubiquitin. Right panel in B and C: Corresponding SDS-PAGE gel stained with Instant Blue™ to visualize the total protein content. The *in vitro* ubiquitination reaction contains: 1 μM Ube1 (E1), 10 μM UbcH5c (E2), 5 μM $HOIP_{697-1072}$, 25 μM N4BP1 construct, 50 μM mono-ubiquitin, 1 μM Cy5-labelled mono-ubiquitin. As a control, there was no N4BP1 added into the reaction. Assay buffer for the reaction is described in detail in section 3.2.4.2. Reaction was initiated by adding ATP at a final concentration of 5 mM. Reaction temperature: 25 °C. M: standard protein ladder. HMW chains: high molecular weight chains.

As mentioned in section 4.1.1, the identified N4BP1-binding region in HOIP is located in the polypeptide of residues 299-438 by the pull-down experiments (Figure 4.1.1E, Figure 4.1.2 B, C). This region is upstream of $HOIP_{697-1072}$. It is also shown that the $N4BP1_{1-392}$ and $HOIP_{697-1072}$ do not interact with each other when probed by ITC (Please see Appendix VI, Figure V6). The binding data support the notion that an allosteric effect in the active site of HOIP takes place upon the interaction with N4BP1. Therefore, a longer construct $HOIP_{300-1072}$ is used in the following studies.

4.1.3.2 N4BP1 UBA-like domain mediates a pronounced inhibitory effect on the E3 ligase activity of $HOIP_{300-1072}$

In addition to the catalytic RBR domain, $HOIP_{300-1072}$ includes the binding region to N4BP1 which was suggested by the pull-down experiments (Figure 4.1.1E and Figure 4.1.2C). Compared to the full-length protein, $HOIP_{300-1072}$ lacks the N-terminal PUB domain (Figure 4.1.2A). The construct requires the presence of the LUBAC subunit HOIL-1L which releases the auto-inhibitory effect from the HOIP UBA domain. Only when HOIL-1L binds to the HOIP UBA domain Ub chains synthesis can be observed.

To this end, all the *in vitro* ubiquitination assays were performed with the addition of HOIL-1L. From the results, $N4BP1_{1-392}$ displays strong inhibition on the E3 ligase activity of $HOIP_{300-1072}$. As seen from the 60-min and 120-min two time points, the high molecular weight linear ubiquitin chain formation is strongly inhibited compared to the control (Figure 4.1.5).

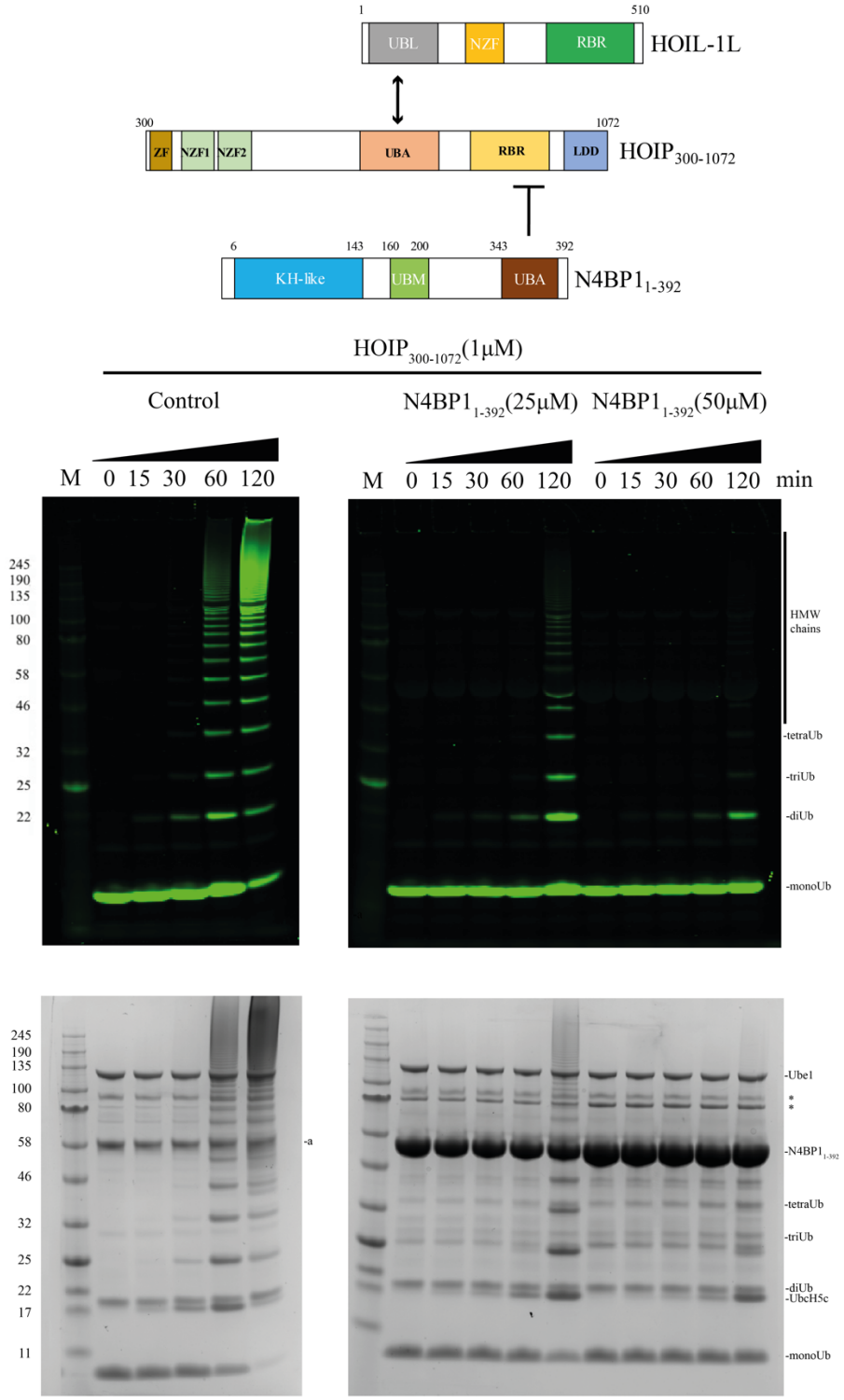


Figure 4.1.5 The inhibitory effect of N4BP1$_{1-392}$ on HOIP$_{300-1072}$. Upper panel: Schematic of the combination of HOIP, HOIL-1L and N4BP1 constructs in the reaction. Middle panel: The results of *in vitro* inhibitory assay of N4BP1$_{1-392}$ on HOIP$_{300-1072}$ at the indicated time points by SDS-PAGE. The gel was scanned at an excitation wavelength of 633 nm and an emission wavelength of 670 nm in order to visualize Cy5-labelled ubiquitin. Bottom panel: Corresponding SDS-PAGE gel stained with Instant Blue™ to visualize the total protein content. The *in vitro* ubiquitination reaction contains: 1 μM Ube1 (E1), 10 μM UbcH5c (E2), 1 μM HOIP$_{300-1072}$, 5 μM HOIL-1L (the protein band is indicated as "a"), 25 μM or 50 μM N4BP1$_{1-392}$ construct, 50 μM mono-ubiquitin, 1 μM Cy5-labelled mono-ubiquitin. As a control, there was no N4BP1 added into the reaction. Assay buffer for the reaction is described in detail in section 3.2.4.2. Reaction was initiated by adding ATP at a final concentration of 10 mM. Reaction temperature: 25 °C. M: standard protein ladder. * indicates the impurity band from purifying HOIL-1L. HMW chains: high molecular weight chains.

In order to identify the minimal region required for inhibition, N4BP1$_{1-392}$ was divided into two parts: 1) the UBA-like domain alone (N4BP1$_{312-392}$); 2) the construct N4BP1$_{1-311}$ which contains both the KH-like domain and the UBM-like domain. The two constructs were employed in the ubiquitination assay respectively.

From the results shown in Figure 4.1.6, the UBA-like domain alone is efficient to inhibit the linear ubiquitin chain formation by HOIP$_{300-1072}$, in contrast to the previous observation where the UBA-like domain does not inhibit HOIP$_{697-1072}$ (Figure 4.1.4B). This might be due to the lack of the N4BP1 binding region in the HOIP$_{697-1072}$ construct. The additional support from the binding between N4BP1$_{1-392}$ and HOIP$_{300-1072}$ may facilitate to place the UBA-like domain in an appropriate position for efficient inhibition on HOIP.

Interestingly, the shorter construct N4BP1$_{1-311}$ still retains weak inhibitory function (Figure 4.1.6). This is in line with the previous observations that the KH-like domain in this construct displays a moderate inhibitory effect on HOIP$_{697-1072}$. This weak inhibition remains to be verified by using the KH-like domain alone for the ubiquitination assay by HOIP$_{300-1072}$.

To conclude, the strong inhibitory effect from N4BP1 UBA-like domain and the weak inhibitory effect from KH-like domain, as well as the key binding regions between N4BP1 and HOIP, work together to facilitate N4BP1 inhibiting on the E3 ligase activity of HOIP. Further structural studies would help to elucidate the exact inhibitory mechanisms from the N-terminal N4BP1 on the ligase activity of HOIP.

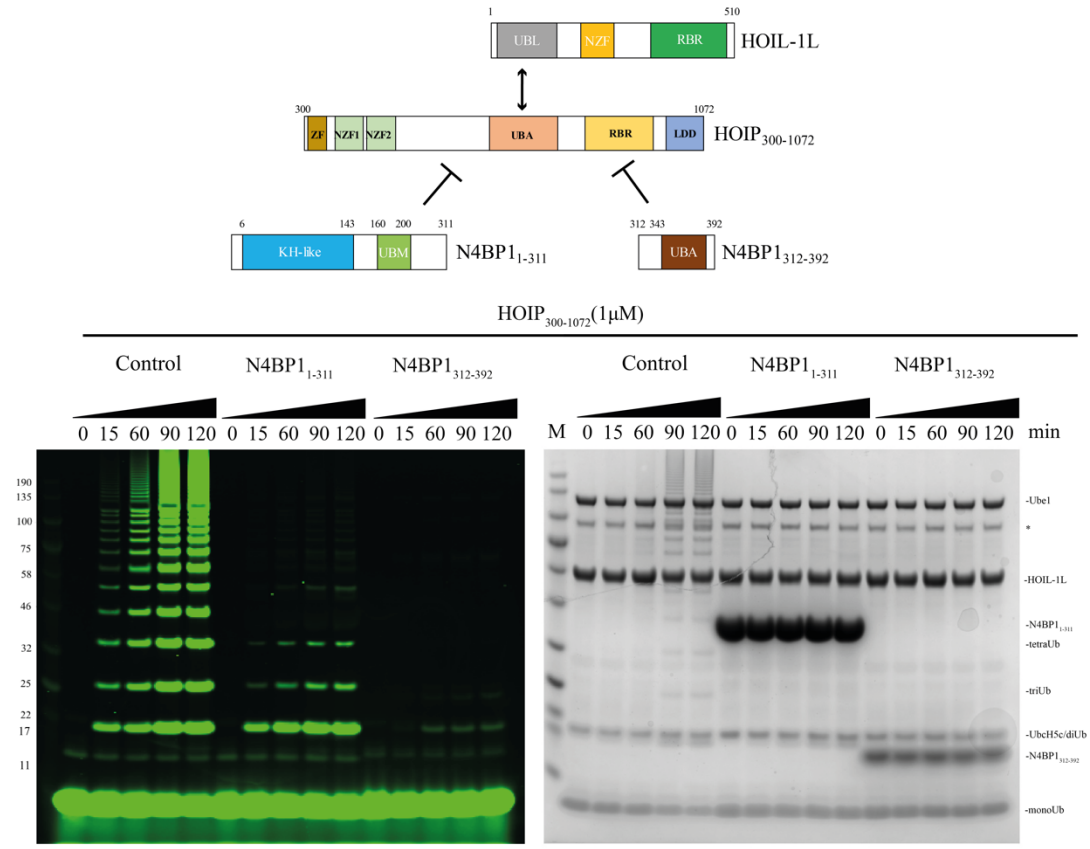


Figure 4.1.6 The inhibitory effect of N4BP1$_{1-311}$ and the UBA-like domain (N4BP1$_{312-392}$) on HOIP$_{300-1072}$. Upper panel: Schematic of the combination of HOIP, HOIL-1L and N4BP1 constructs in the reaction. Bottom panel: The results of *in vitro*

inhibitory assay of each N4BP1 construct on $HOIP_{300-1072}$ at the indicated time points by SDS-PAGE. The gel was scanned at an excitation wavelength of 633 nm and an emission wavelength of 670 nm in order to visualize Cy5-labelled ubiquitin. Corresponding SDS-PAGE gel stained with Instant Blue™ was shown on the right to visualize the total protein content. The *in vitro* ubiquitination reaction contains: 1 μM Ube1 (E1), 10 μM UbcH5c (E2), 1 μM $HOIP_{300-1072}$, 5 μM HOIL-1L (the protein band is indicated as "a"), 50 μM N4BP1 construct, 100 μM mono-ubiquitin, 1 μM Cy5-labelled mono-ubiquitin. As a control, there was no N4BP1 added into the reaction. Assay buffer for the reaction is described in detail in section 3.2.4.2. Reaction was initiated by adding ATP at a final concentration of 10 mM. Reaction temperature: 25 °C. M: standard protein ladder. * indicates the impurity band from purifying HOIL-1L.

4.1.3.3 Crystallization of $N4BP1_{1-392}$

After elucidating the inhibitory function of the construct $N4BP1_{1-392}$, it was decided to investigate the molecular structure by X-ray crystallography. The protein was successfully purified with high purity and quality via GSH affinity chromatography and size-exclusion chromatography (Figure 4.1.7). Using the LMB screen kit (Molecular Dimension), several crystal candidates can be obtained but unfortunately no diffraction pattern was observed when the crystals were exposed to X-rays at the Diamond light source.

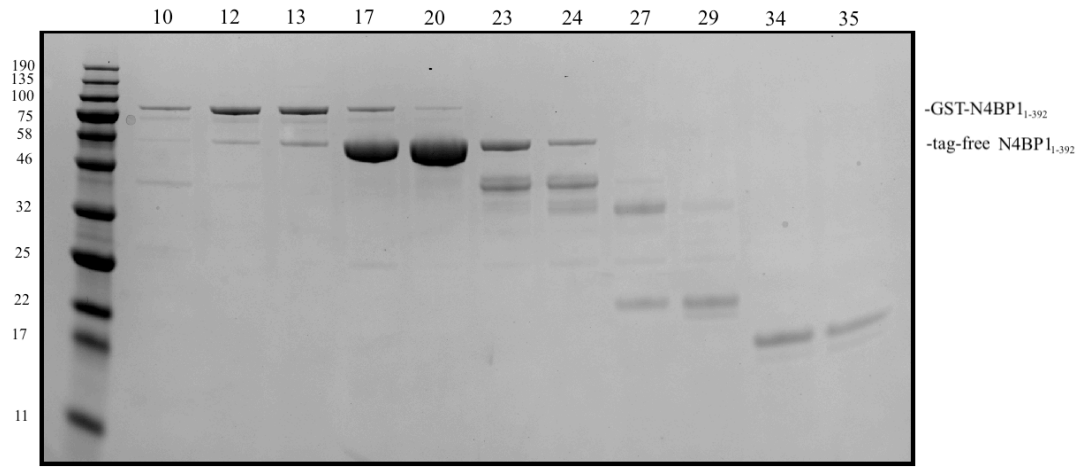


Figure 4.1.7 The purification of N4BP1$_{1-392}$ by size-exclusion chromatography (SEC). Protein from GSH affinity chromatography was incubated with HRV-3C protease overnight. The sample were then subjected to size-exclusion chromatography as shown here. Fraction 20 was used for crystallography.

Due to the important role of the N4BP1 UBA-like domain for the inhibition of HOIP, a structural modeling was carried out to further provide some molecular insights into the 3-dimensional organization of this domain. Using Phyre2 for a secondary structural prediction reveals that the N4BP1 UBA-like domain displays an ordered structure for residues 342-385 which features a 3-helix fold (Figure 4.1.8A). The top template for the prediction is the NTD domain from MCPIP1 (PDB: 2N5J). The N4BP1 UBA-like domain has a 53% coverage with this template at a confidence score of 96.6%. The *de novo* prediction by trRosetta also gave similar results (Figure 4.1.8B).

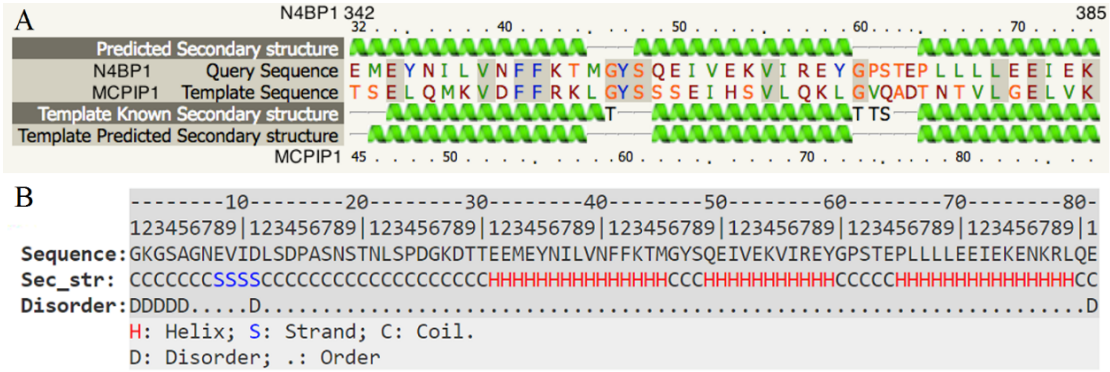


Figure 4.1.8 The predicted secondary structure of N4BP1 UBA-like domain by Phyre2 (A) and trRosetta (B). The structured region is indicated in green helices or "H". The MCPIP1 NTD domain (PDB: 2N5J) is the template for the prediction by Phyre2. The protein sequence submitted to trRosetta is N4BP1$_{312-392}$.

The structure model of the N4BP1 UBA-like domain by Phyre2 is based on the MCPIP1 NTD domain as a template, it is no surprise that the model is highly similar to the template (Figure 4.1.9A). The *de novo* modelling by trRosetta generated five models with low energy which are also similar to the

MCPIP1 NTD domain. The only difference is that the trRosetta model has defined several extra residues in both the N-terminus and C-terminus of this domain as structured helices, resulting in longer α1 and α3 helices than the MCPIP1 NTD domain (Figure 4.1.9B). Both models support the conservation between the N4BP1 UBA-like domain and the MCPIP1 NTD domain.

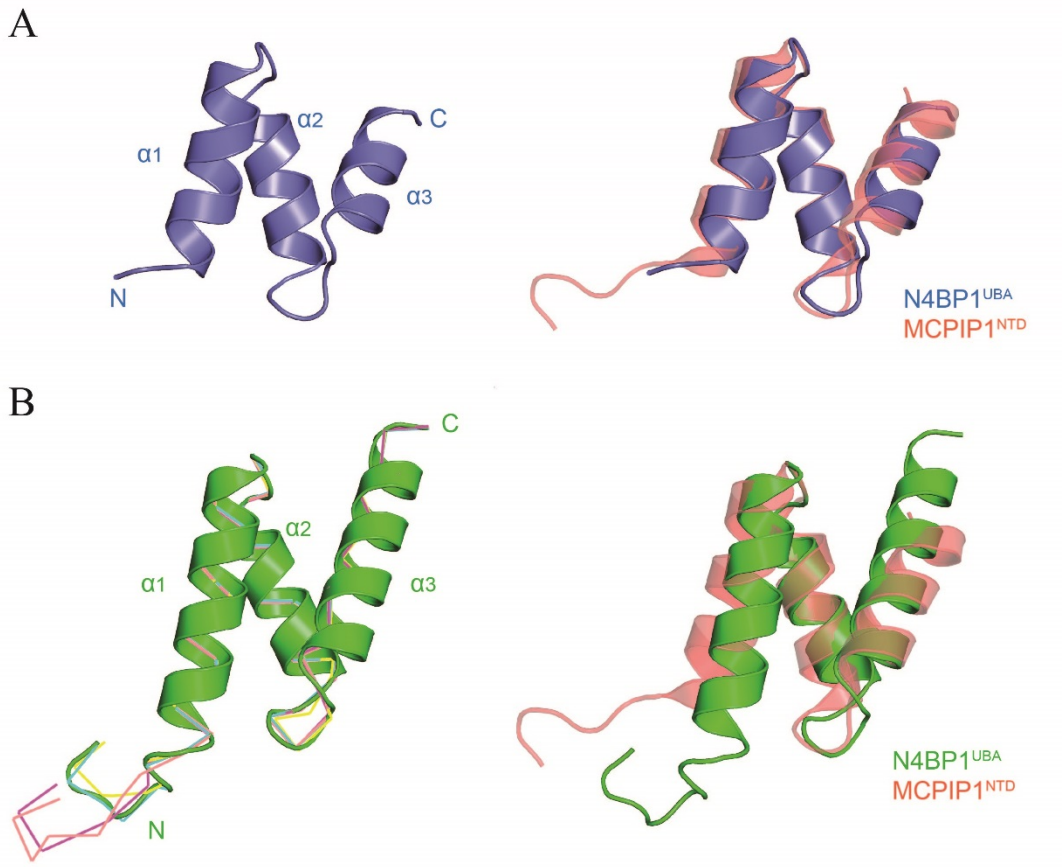


Figure 4.1.9 The structure models of N4BP1 UBA-like domain by Phyre2 (A) and trRosetta (B). The model by Phyre2 is shown in blue. Five models with low energy are generated by trRosetta. The top model with the lowest energy is shown as cartoon in green while the rest of the four models are shown as ribbons. The structure comparison between the N4BP1 model and the MCPIP1 NTD domain (PDB: 2N5J) is displayed on the right in each figure. The top model by trRosetta is used for structure comparison.

4.1.4 Discussion

In this chapter, the inhibitory role of N-terminal N4BP1 ($N4BP1_{1-392}$) in LUBAC activity was studied. It was found that the N4BP1 UBA-like domain has a strong inhibition on the E3 ligase activity of the longer HOIP construct ($HOIP_{300-1072}$), but does not have the inhibition on the short HOIP construct ($HOIP_{697-1072}$), indicating that the binding region between N4BP1 and HOIP is important to facilitate the inhibition. Noteworthy, HOIL-1L is required to release the auto-inhibition of $HOIP_{300-1072}$ by binding to the HOIP UBA domain. However, $N4BP1_{1-392}$ does not bind to HOIL-1L (Please see Appendix VI, Figure V7), which indicates that N4BP1 and HOIP do not compete to interact with HOIL-1L. Thus, it excludes the possibility that N4BP1 prevents the release of HOIP auto-inhibition via binding to HOIL-1L. Apart from the UBA-like domain, the KH-like domain also has a moderate inhibition on HOIP but does not require the binding between N4BP1 and HOIP. Taken together, the inhibitory effect from both domains together with the key binding regions between N4BP1 and HOIP maximizes the inhibition from N4BP1 on LUBAC.

Structural studies would assist to reveal the exact inhibitory mechanisms from N4BP1. To this end, the trials to crystallize $N4BP1_{1-392}$ was attempted but unfortunately not successful. For the future work, the crystallization conditions will be optimized, hopefully allowing us to obtain protein crystals with high quality. The trials to crystallize the N4BP1/HOIP complex is also important to ultimately elucidate the inhibition mechanism from the structural perspectives.

The structure modelling of the N4BP1 UBA-like domain by both Phyre2 and trRosetta indicates that the N4BP1 UBA-like domain is structurally conserved with the MCPIP1 NTD domain. It is reported that the presence of MCPIP1 NTD domain facilitates the RNAse activity of the RNAse domain from the same protein. Structural study reveals that the interaction between

the two domains hinders the oligomerization of the RNAse domain and keeps the RNAse domain in the functional dimer states (See the short review in chapter 1.4.3) (Yokogawa et al., 2016). In addition, the RNAse domain within the C-terminal part of N4BP1 is also predicted to have a similar structure to the MCPIP1 RNAse domain (this will be introduced in detail in section 4.2.2.5.1). To gain a functional insight, it would be interesting to investigate the possible interaction between the UBA-like domain and the RNase domain in N4BP1.

It is known that MCPIP1 is involved in inflammatory signaling and plays as a negative regulator in NF-κB activation. These structural and functional similarities between N4BP1 and MCPIP1 further support the importance of N4BP1 involved in the TNFR signaling pathway. Since the structure of each domain of MCPIP1 has been solved (Yokogawa et al., 2016), it can be used as a starting point for the structure investigation of N4BP1 in the future study.

Chapter 4.2

Quantitative analysis of the N4BP1 ubiquitin-binding properties

4.2.1 Introduction: The N4BP1 CUE domain

The C-terminus of N4BP1 (393-892) contains two defined domains, the RNAse domain and the CUE domain (Figure 4.2.1). Bioinformatic analysis has grouped the N4BP1 CUE domain into the canonical CUE domain family, which is known to be associated with ubiquitin binding due to its key conserved motifs.

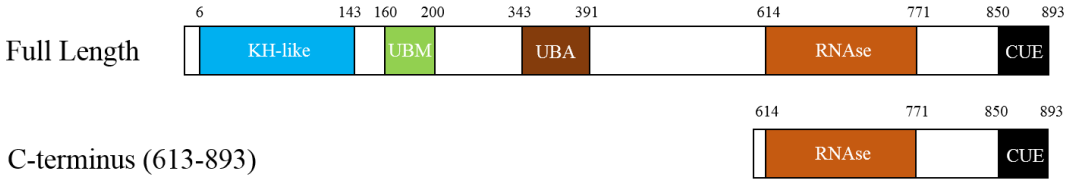


Figure 4.2.1 The domain structures of the C-terminus and full length N4BP1.

As introduced in section 1.4.1, N4BP1 is identified as a linear ubiquitin receptor. Based on binding studies by pull-down experiments, the CUE domain of N4BP1 shows an apparent binding preference to di-ubiquitin chains of different linkage types with the exception K48-diUb and mono-ubiquitin (Figure 4.2.2, GST-mN4BP1 (850-893)). Furthermore, if the construct is extended to include part of the RNAse domain (Figure 4.2.2, GST-mN4BP1 (706-893)), the binding specificity is restricted to M1-diUb and K63-diUb chains (Kliza et al., unpublished data). This may indicate that to some extent the RNAse domain is also involved in ubiquitin binding and may help enhance the binding specificity to M1-diUb and K63-diUb chains.

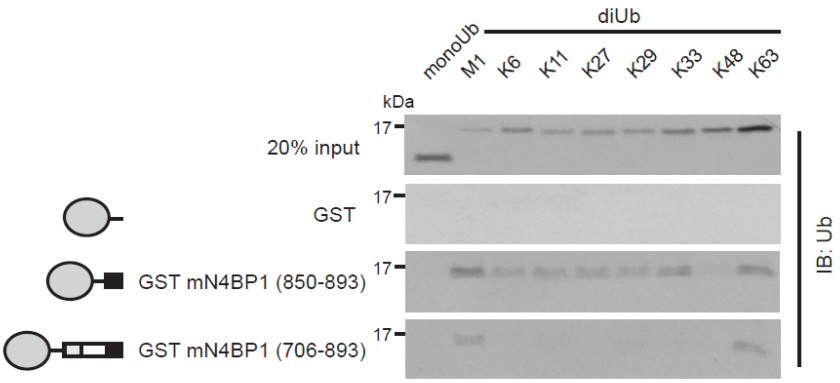


Figure 4.2.2 Pull-down experiments of the N4BP1 CUE domain (850-893) or extended CUE domain (706-893) with mono-ubiquitin and different types of di-ubiquitin chains (Kliza et al., unpublished data).

In this chapter, an in-depth investigation of the ubiquitin-binding properties of the C-terminal domain of N4BP1 will be performed. Section 4.2.2.1 to 4.2.2.4 are centred on the interaction between the CUE domain and ubiquitin. The study on RNAse domain will be discussed in section 4.2.2.5.

Section 4.2.2.1 is focused on the quantitative characterization of the ubiquitin-binding properties of the CUE domain by SPR and ITC. However, resolving the ubiquitin-binding mode requires structural approaches. Hence, the structural description of the CUE domain by solution NMR is presented in section 4.2.2.2. This structure further allows us to study the interaction between the CUE domain and ubiquitin from the single-residue level by NMR titrations. This will be described in section 4.2.2.3. Taking all the above data together, a CUE/ubiquitin complex model is generated in section 4.2.2.4.

Section 4.2.2.5 is focused on the binding specificity of N4BP1 to ubiquitin mediated by the RNAse domain. A model is proposed in section 4.2.3.3 to help explain how the CUE domain and RNAse domain work together to facilitate the ubiquitin-binding specificity.

4.2.2 Results

4.2.2.1 Quantitative analysis of the interaction between the N4BP1 CUE domain and ubiquitin probed by ITC and SPR

4.2.2.1.1 The N4BP1 CUE domain specifically binds to ubiquitin but not NEDD8

The canonical CUE domain was initially characterized by two major conserved motifs required for ubiquitin binding: an FP motif and a di-leucine motif (Shih et al., 2003; Figure 4.2.3). Sequence alignments of the mouse N4BP1 CUE domain with other CUE domain family members highlight the level of conservation of the FP motif. The N4BP1 CUE domain does not have the strictly conserved di-leucine motif, but evidence from gp78 CUE domain where an IL motif is present has demonstrated that this motif is involved in ubiquitin binding (Liu et al., 2012).

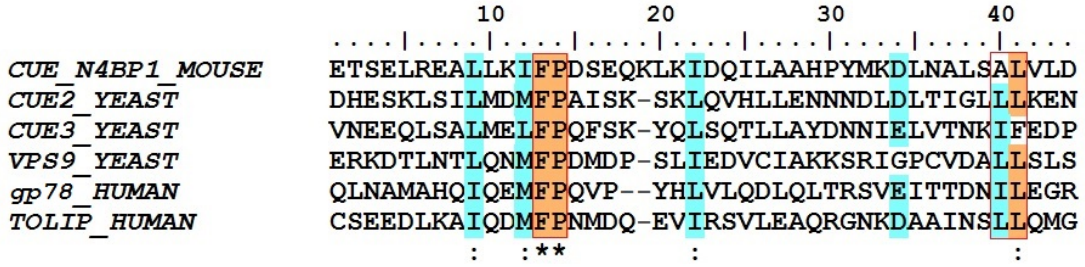


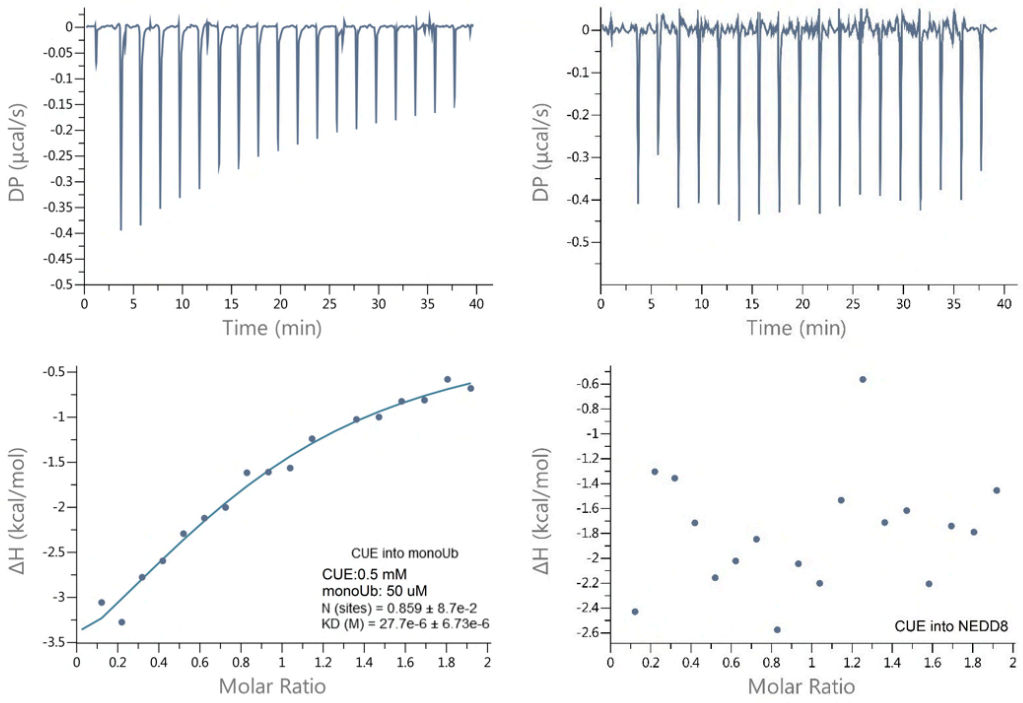
Figure 4.2.3 Sequence alignments of the mouse N4BP1 CUE domain with the CUE domain family members. Identical residues are shaded in orange. The residues with >80% similarities are shaded in cyan. The FP motif and di-leucine motif are highlighted with red box. Sequence alignments were performed by ClustalW and visualized by bioedit.

Based on an analysis by the protein fold recognition server Phyre2, the structure of N4BP1 CUE domain can be predicted using the CUBAN domain from KHNYN (PDB: 2N5M) as the top template. The N4BP1 CUE domain displays 93% template coverage with a confidence score of 99.6%. The KHNYN CUBAN domain is consisted of a three-helix bundle arrangement, while the protein fold is different from the classical CUE domain. It was

shown that the protein has a binding preference to the ubiquitin-like protein NEDD8 instead of ubiquitin (See section 1.4.2; Castagnoli et al., 2019).

In order to quantify the interaction between the N4BP1 CUE domain (850-893) and ubiquitin observed by pull-down experiments (Figure 4.2.2), ITC measurements were performed, which allow precise determination of the affinity between the two proteins (Figure 4.2.4).

A



B

Protein pairs	Kd (uM)	N(sites)
CUE-like domain (tag-free): monoUb	27.7 ± 6.73	0.86 ± 0.09
CUE-like domain (tag-free): NEDD8	undetectable	undetectable

Figure 4.2.4 ITC measurements and the parameter summary of titrating the N4BP1 CUE domain into mono-ubiquitin and NEDD8 respectively. Each ITC plot contains the row data (upper) and the binding isotherm (lower) for the interactions. The data for CUE: monoUb were fitted to a single-site model using the MicroCal PEAQ-ITC Analysis Software following the method described in section 3.2.5.2. Measurement conditions: Syringe: 0.5 mM CUE, Cell: 50 uM monoUb or NEDD8. 18 injections, 2 µL/injection, duration: 2s, spacing time: 120 s, measurement temperature: 25℃.

The tag-free CUE domain was used for the ITC experiments. Despite the high sequence similarity between the KHNYN CUBAN domain and the N4BP1 CUE domain, the interaction between the N4BP1 CUE domain and NEDD8 is undetectable by ITC. However, in contrast to the pull-down experiments where the GST-tagged protein was used for the binding test, the tag-free N4BP1 CUE domain shows clear binding to mono-ubiquitin with a Kd value of 27.4 μM from the ITC measurements (Figure 4.2.4). The differences of the ubiquitin-binding properties from the GST-tagged and the tag-free N4BP1 CUE domain will be discussed in the following content.

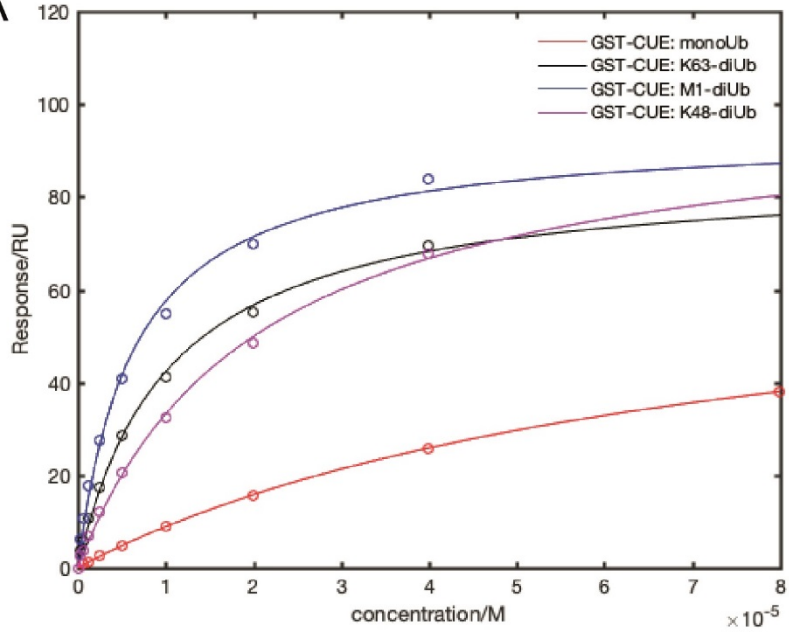
4.2.2.1.2 The interaction between the N4BP1 CUE domain and ubiquitin probed by SPR

To confirm the pull-down results of the N4BP1 CUE domain and ubiquitin (Figure 4.2.2), SPR measurements were performed to quantitatively characterize the binding affinity of N4BP1 CUE domain to mono-ubiquitin (shorten to monoUb hereafter) and di-ubiquitin chains (shorten to M1/K48/K63-diUb respectively hereafter). The GST-tagged CUE domain (shorten to GST-CUE hereafter) was immobilized on an anti-GST antibody surface of a CM5 chip. MonoUb or diUb chains were in the mobile phase and running separately through the chip surface.

The binding constant was calculated using the amplitude of the SPR signal since it was not possible to determine the very fast association and dissociation rates under these conditions. The response data points at the equilibrium stage from the sensorgrams (raw data not shown) were plotted against the ubiquitin concentration. The data were fitted to obtain the equilibrium dissociation constant (K_D) as shown in Figure 4.2.5 (Please see section 3.2.5.3 for more data fitting details). The SPR results demonstrate the high affinity of M1-diUb to GST-CUE with a K_D value of 6.38 μM, which is 11-fold stronger than the monoUb, suggesting that GST-CUE specifically recognizes ubiquitin chains. However, this specificity is not exclusive to

ubiquitin chains with linear linkages. K63-diUb also shows a high affinity to GST-CUE with a K_D value of 10.31 μM, which is comparable to M1-diUb. K48-diUb binds to GST-CUE with a K_D value of 20.42 μM, indicating a less affinity than M1- and K63-diUb, but still higher than monoUb. Altogether, these SPR results are in line with the pull-down assay from Figure 4.2.2.

A



B

Analytes	Kd (μM)
monoUb	71.23±0.84
M1-diUb	6.38±0.63
K48-diUb	20.42±1.01
K63-diUb	10.31±0.64

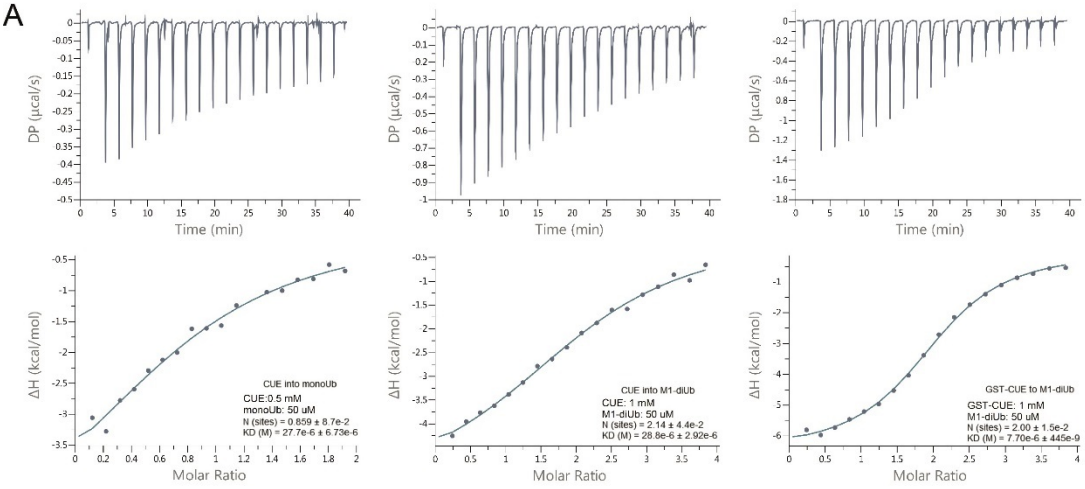
Figure 4.2.5 Quantitative binding measurements of GST-tagged N4BP1 CUE domain to monoUb and M1/K48/K63-diUb by SPR. A. The response data points at the equilibrium stage derived from the sensorgrams plotted against the ubiquitin concentration. B. Parameter table of K_D summarized from the plot. Data were fitted by Matlab following the method described in section 3.2.5.3.

4.2.2.1.3 The interaction between the N4BP1 CUE domain and ubiquitin probed by ITC

1) GST tag artificially increases the affinity of the CUE domain to M1-diUb

The above SPR measurements have shown that the GST-tagged CUE domain has ubiquitin linkage preference. It is also interesting to determine the stoichiometry of these interactions. To this end, ITC measurements are ideally suited to determine the number of binding sites in solution. The tag-free CUE domain was used for the ITC experiments (Figure 4.2.6).

A



B

Protein pairs	Kd (µM)	N(sites)
CUE (tag-free): monoUb	27.7 ± 6.73	0.86 ± 0.09
CUE (tag-free): M1-diUb	28.8 ± 2.92	2.14 ± 0.04
CUE (GST-tagged): M1-diUb	7.70 ± 0.45	2.00 ± 0.02

Figure 4.2.6 Quantitative binding measurements of the N4BP1 CUE domain to monoUb and M1-diUb by ITC. A. ITC titration isotherms. B. Parameter table of Kd summarized from A. Data were fitted to a single-site model using the MicroCal PEAQ-ITC Analysis Software following the method described in section 3.2.5.2. Measurement conditions: Syringe: 0.5 mM or 1 mM CUE (specified in each isotherm), Cell: 50 uM monoUb or M1-diUb. 18 injections, 2 µL/injection, duration: 2s, spacing time: 120 s, measurement temperature: 25℃.

From the ITC results, the tag-free CUE domain shows very similar affinity to both monoUb and M1-diUb with a Kd value of ~ 28 μM. The CUE domain binds to monoUb in a 1:1 ratio and binds to M1-diUb in a 2:1 ratio. It most likely indicates that each ubiquitin moiety in M1-diUb interacts with one molecule of the CUE domain. Surprisingly, the Kd values obtained from ITC are different from those derived by SPR. Particularly, the affinities of the CUE domain to monoUb and M1-diUb are not similar to each other in the SPR measurements where CUE domain seems to prefer M1-diUb over monoUb and its binding to monoUb is very weak (Figure 4.2.5). Since GST-CUE was used for the SPR measurements, the GST tag might interfere with the interaction. To verify this hypothesis, the binding affinity of GST-CUE to M1-diUb was measured by ITC. As seen in Figure 4.2.6, the result from ITC (7.70 μM) is very similar to that from SPR (6.38 μM), both demonstrating that the GST-CUE shows much higher affinity to M1-diUb compared to the tag-free protein.

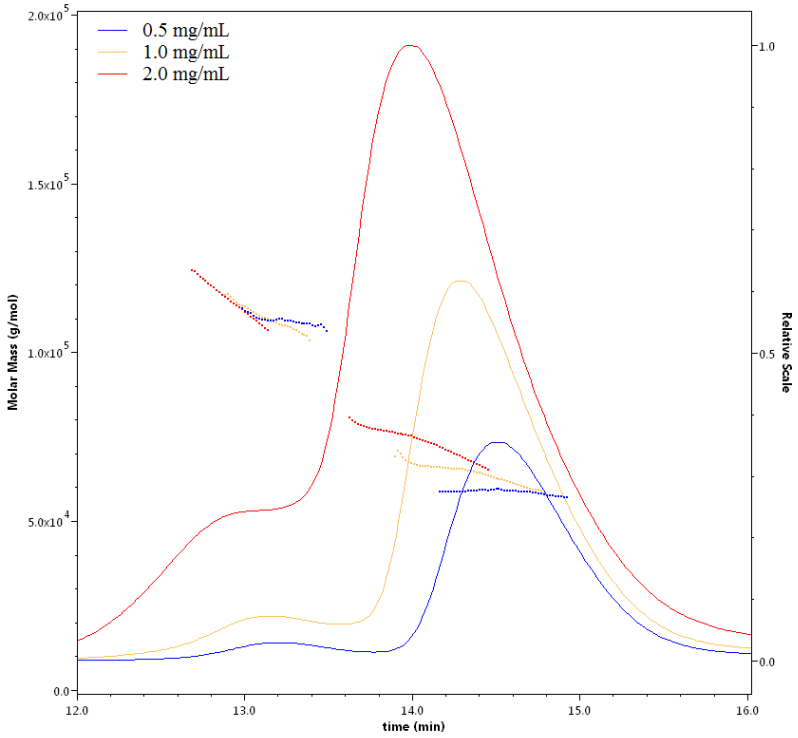


Figure 4.2.7 SEC-MALLS measurements of GST-tagged CUE domain. Three concentrations of GST-CUE at 0.5, 1.0 and 2.0 mg/mL were used respectively.

Due to the dimerization effect from the GST, the oligomeric state of GST-CUE was measured by SEC-MALLS. GST-CUE monomer is ~ 32 KDa (3.2×10^4 g/mol). To our expectation, the GST-CUE species are mostly in dimeric states (Molar mass at ~ $5.5\text{-}8.0\times10^4$ g/mol) at three different concentrations (Figure 4.2.7). This result suggests that the GST dimer could artificially bring the two CUE domains in close proximity. Therefore, the affinity between the GST-CUE and M1-diUb is increased.

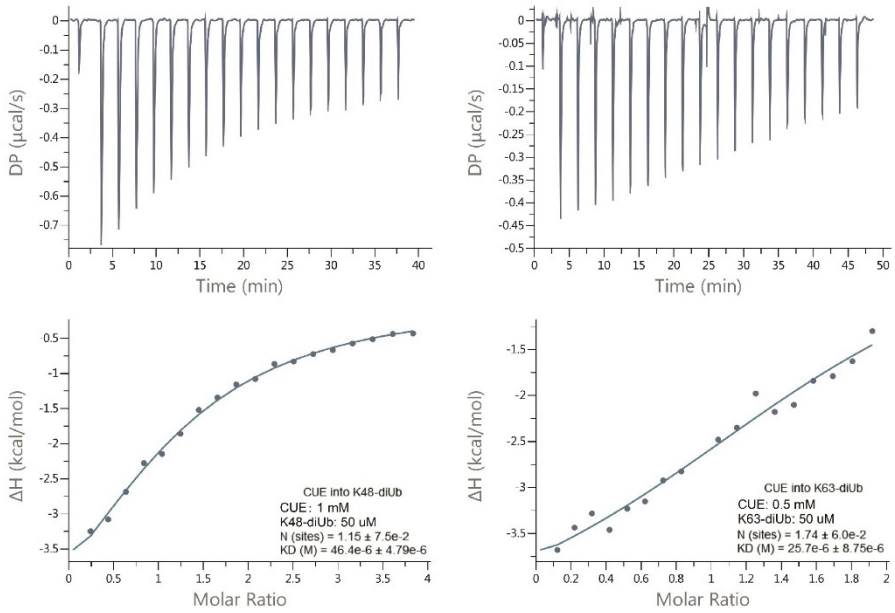
2) The N4BP1 CUE domain distinguishes K48-diUb from monoUb, M1-diUb and K63-diUb

The binding between the tag-free CUE domain with K48- and K63-diUb chains was also measured by ITC respectively (Figure 4.2.8). Together with the above ITC results, it indicates that the CUE domain binds to monoUb, M1-diUb and K63-diUb at a comparable level (with a Kd value of 26~29 μM), while its affinity to K48-diUb is the weakest with a Kd value of 46.6 μM.

In addition, compared to the ~2:1 ratio between the CUE domain and M1-diUb or K63-diUb, the binding stoichiometry between the CUE domain and K48-diUb is ~1:1. This indicates that one K48-diUb molecule could only bind to equal amount of the CUE domain.

To conclude, among monoUb and three diUb chains used here, the N4BP1 CUE domain could distinguish K48-diUb from monoUb, M1-diUb and K63-diUb.

A



B

Protein pairs	Kd (µM)	N(sites)
CUE (tag-free): K48-diUb	46.6 ± 4.79	1.15 ± 0.08
CUE (tag-free): K63-diUb	25.7 ± 8.75	1.74 ± 0.06

Figure 4.2.8 Quantitative binding measurements of the N4BP1 CUE domain to K48-diUb and K63-diUb by ITC. A. ITC titration isotherms. B. Parameter table of Kd summarized from A. Data were fitted to a single-site model using the MicroCal PEAQ-ITC Analysis Software following the method described in section 3.2.5.2. Measurement conditions: Syringe: 1 mM CUE, Cell: 50 uM K48-diUb or K63-diUb. 18 injections, 2 µL/injection, duration: 2s, spacing time: 120 s, measurement temperature: 25℃.

In the next section, the structure model of the N4BP1 CUE domain was solved by solution NMR. This will help characterize the ubiquitin-binding properties of the CUE domain by the means of NMR titrations in section 4.2.2.3.

4.2.2.2 NMR solution structure model of the N4BP1 CUE domain

NMR spectroscopy was chosen as the method of choice for the determination of the structure of the N4BP1 CUE domain for two reasons. First of all, the size of the CUE domain is only ~5 KDa, ideal for solving the structure by NMR. Secondly, the following NMR titrations are suited to study the dynamics of the interaction between the CUE domain and ubiquitin.

The CUE domain was labelled with ^{13}C and ^{15}N and purified to a very high standard required for NMR. The backbone structure was determined by solution NMR using the 3-dimensional experiments including HSQC, HNCA, NHCOCA, CBCACONH and HNCACB etc. The side chains were assigned using HBHA(CO)NH and HCCH-TOCSY. The structure was modelled by CS-ROSETTA using the experimental chemical shift data collected from the above spectra.

4.2.2.2.1 Purification of the CUE domain

The His-tagged CUE domain was expressed in *E.coli* which was grown in minimal media containing ^{13}C-glucose and $^{15}NH_4Cl$. The His-tagged protein can be successfully enriched via Ni-affinity chromatography. The target protein was subsequently cleaved by on-column cleavage by incubating with HRV-3C protease overnight. The tag-free CUE domain was released from the resin into the buffer and the protein was further purified by size-exclusion chromatography (SEC). Figure 4.2.9 indicates the high purity and quantity of the final protein product that is ready for NMR measurements.

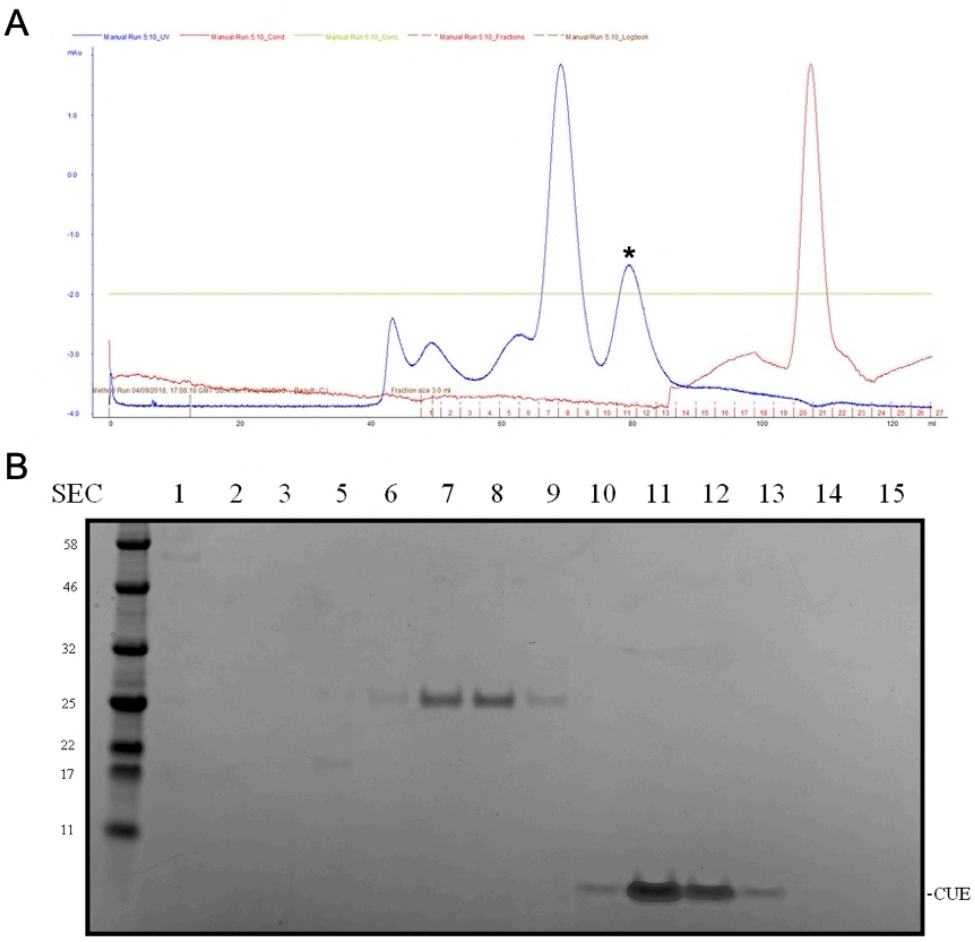


Figure 4.2.9 Purification of $^{15}N/^{13}C$-labelled N4BP1 CUE domain by size-exclusion chromatography (SEC). A. Purification profile by SEC. * indicates the target peak for CUE. B. SDS-PAGE analysis of corresponding peak fractions from A. Fraction 10-13 contain the final pure protein product.

4.2.1.2.2 The stability test of the CUE domain at 37°C

To collect 2- and 3-dimensional NMR spectra with high quality and to minimize measurement time, all spectra were recorded at 37°C. Before recording the full spectra, the stability of the CUE domain at 37°C was tested. A freshly made protein sample was measured at 37°C at different time points (0h, 22h, 69h) to obtain the 1D ^{1}H-NMR signals. The three spectra are almost identical by comparison, including the backbone H^N, backbone Hα and sidechain methyl groups, indicating that the CUE domain is still stable after 69 h at 37°C (Figure 4.2.10).

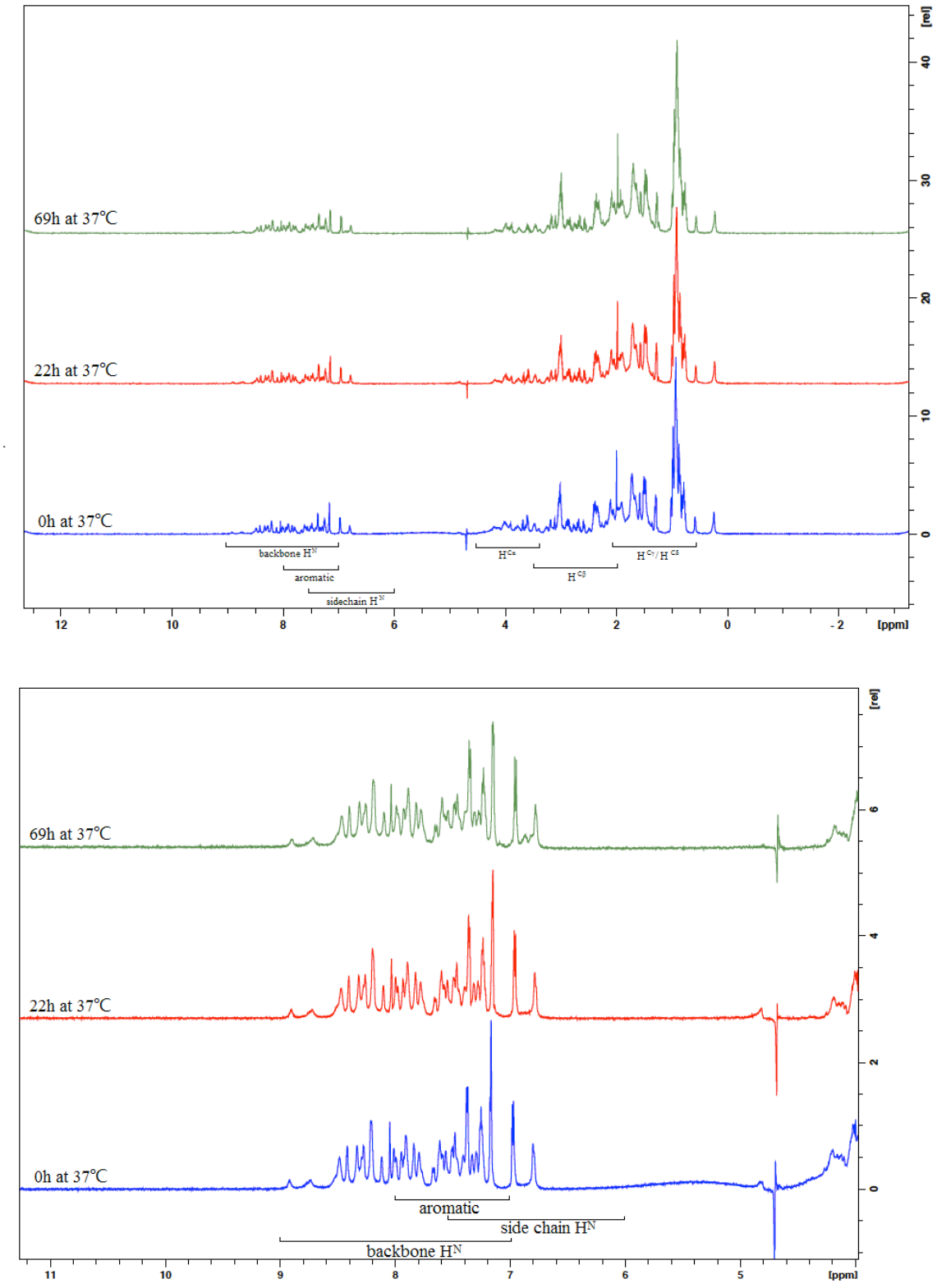


Figure 4.2.10 1D ^{1}H-NMR spectra of the N4BP1 CUE domain measured at three time points at 37°C. Upper panel: Full 1D ^{1}H-NMR spectra. Lower panel: Highlighted H^N area from the full 1D ^{1}H-NMR spectra. Figure was generated by TopSpin.

Furthermore, 2D $^{1}H/^{15}N$-HSQC spectra (SOFAST) were recorded at different temperatures (25°C, 30°C, 34°C, 37°C). In each spectrum, almost all the peaks can be separated from each other. Since each peak in $^{1}H/^{15}N$-HSQC represents a residue (excluding those peaks for the side chains), the results indicate that it is practical to assign the protein with the help from other advanced NMR experiments. By overlapping and comparing all the $^{1}H/^{15}N$-HSQC spectra together, each peak was shifting in an expected order which allows precisely tracking the dynamics of each residue (Figure 4.2.11).

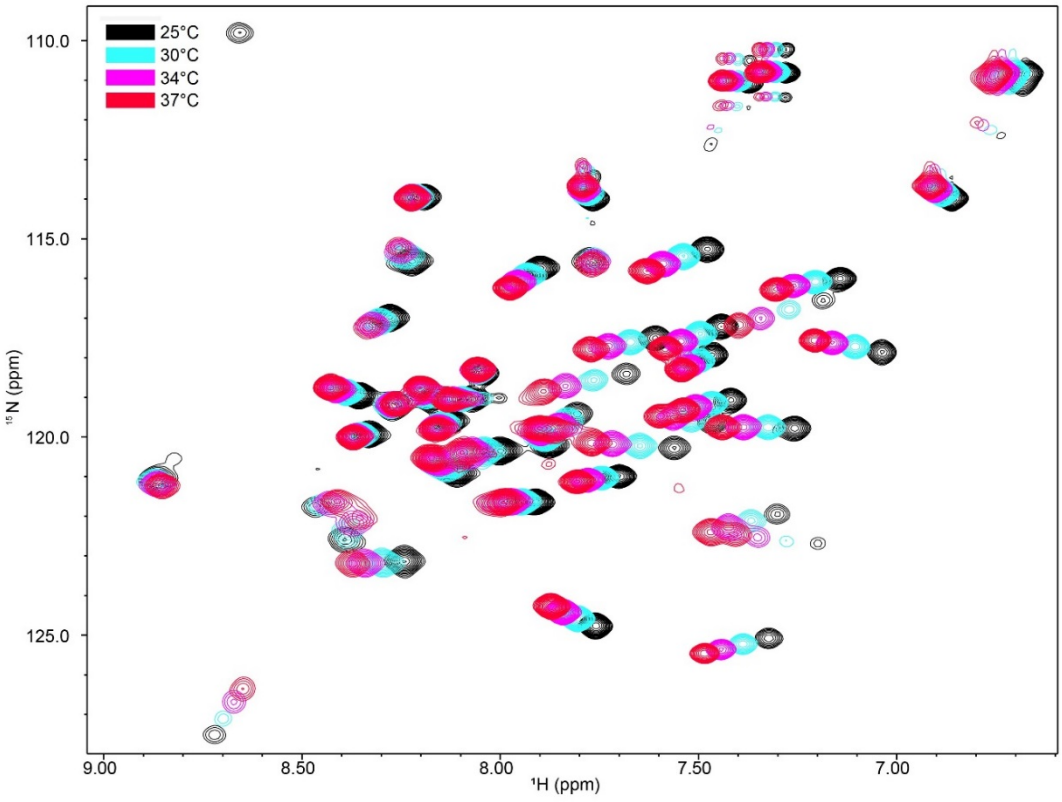


Figure 4.2.11 2D NMR $^{1}N/^{15}N$-HSQC spectra of the N4BP1 CUE domain measured at a temperature range from 25°C to 37°C. Figure was generated by NMRviewJ software.

In conclusion, the N4BP1 CUE domain displays high stability at 37°C for about 3 days, thus recording 3D NMR spectra with high quality and resolution is possible.

4.2.1.2.3 Backbone assignments of the CUE domain

To assign the backbone Cα and carbonyl C (short as CO here) atom of the N4BP1 CUE domain, the tag-free $^{15}N/^{13}C$-double labelled protein was subjected to different 3D NMR measurement combinations, including HN(CA)CO/HNCO, CBCA(CO)NH/HNCACB and HN(CO)CA/HNCA.

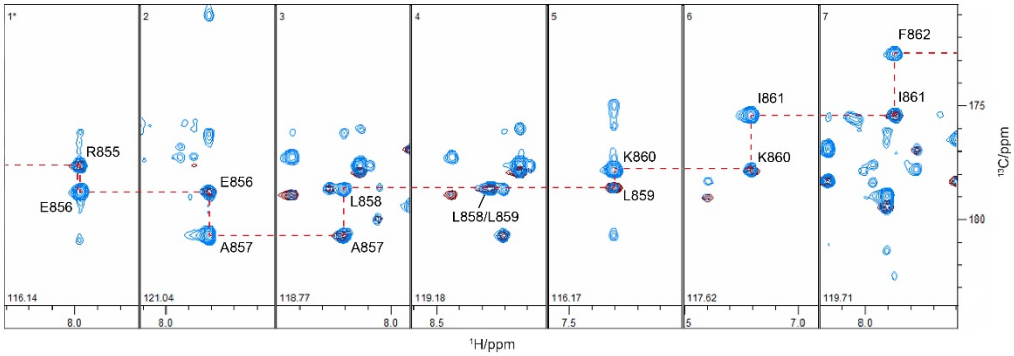
The NMR assignment theory is briefly described in Chapter II. Here is an example of the assignments of residue R855-F862 at each corresponding unique NH strip (Figure 4.2.12).

According to the signature $^{1}N/^{15}N$-HSQC spectrum of the protein, the backbone amino group (N^H and H^N) can be identified but cannot be assigned yet. Each peak from $^{1}N/^{15}N$-HSQC requires a corresponding strip in the 3D spectra which contain not only the chemical shifts of backbone N^H and H^N but also information about the backbone Cα and CO. The backbone assignment is usually conducted in this way.

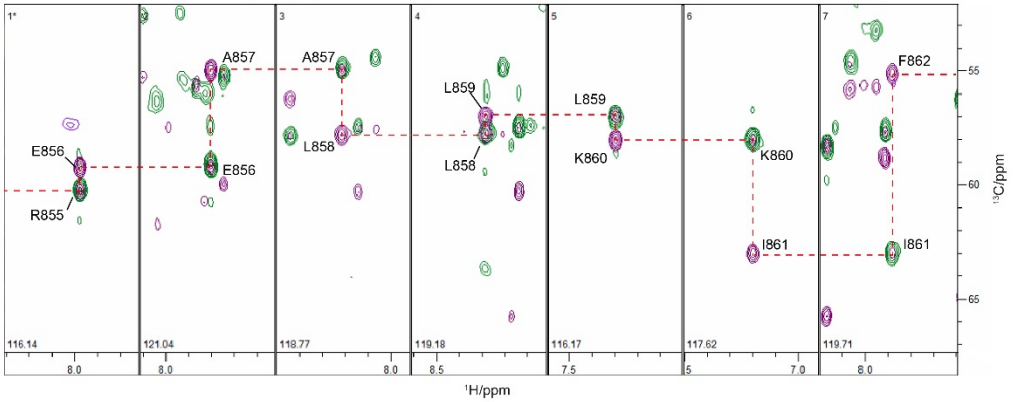
Cαs are usually at a range of 50-70 ppm where their assignments were identified in the first place by HNCA/HN(CO)CA and subsequently confirmed by CBCA(CO)NH/HNCACB. The CBCA(CO)NH/HNCACB experiments also detected the signals from Cβs, usually in a much lower range (18-45 ppm). This allows the discrimination of Cβ signals from Cα signals in the same strip of CBCA(CO)NH/HNCACB. Carbon atoms from the backbone CO group were identified by HNCO/HN(CA)CO.

As can be seen from all the 3 spectra combinations (Figure 4.2.12), the peak of each atom from a certain residue links very well to the next and the preceding ones.

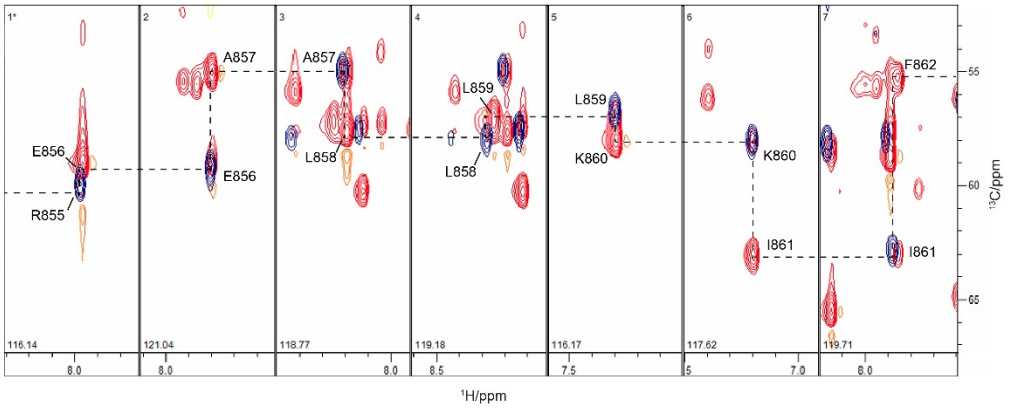
CO assignment-HNCO/HN(CA)CO



Cα assignment-HNCA/HN(CO)CA



Cα assignment-CBCA(CO)NH/HNCACB



Cβ assignment-CBCA(CO)NH/HNCACB

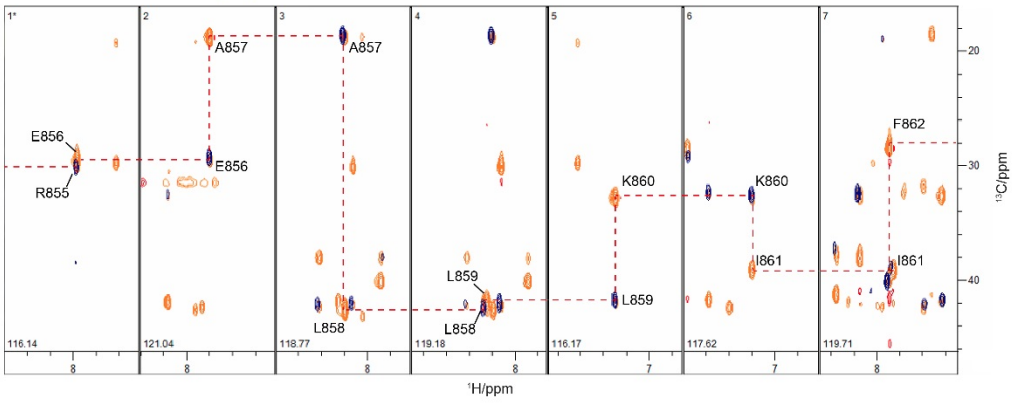


Figure 4.2.12 An assignment example of NH strips of residue R855-F862 by different 3D NMR spectra combinations. Each stripe is in a certain backbone N^H plane of the corresponding residue. The N^H chemical shift value is indicated in the left corner of each strip. The x axis indicates the H^N dimension while the y axis indicates the C dimension. The links between adjacent signals are indicated in dashed lines. Figures were generated by CCPNMR software.

After creating an entire set of strips each individual residue was identified and assigned. Figure 4.2.13 summarizes the final $^1H/^{15}N$-HSQC assignments of almost all the residues except S865 and K882 due to undetectable signals. Note that signals from prolines (P863 and P879) are also absent from $^1H/^{15}N$-HSQC due to the lack of a hydrogen atom attached to the backbone amide. The first 3 residues (GPG) from the tag attached to the N-terminus of the protein are highly flexible and thus not assigned.

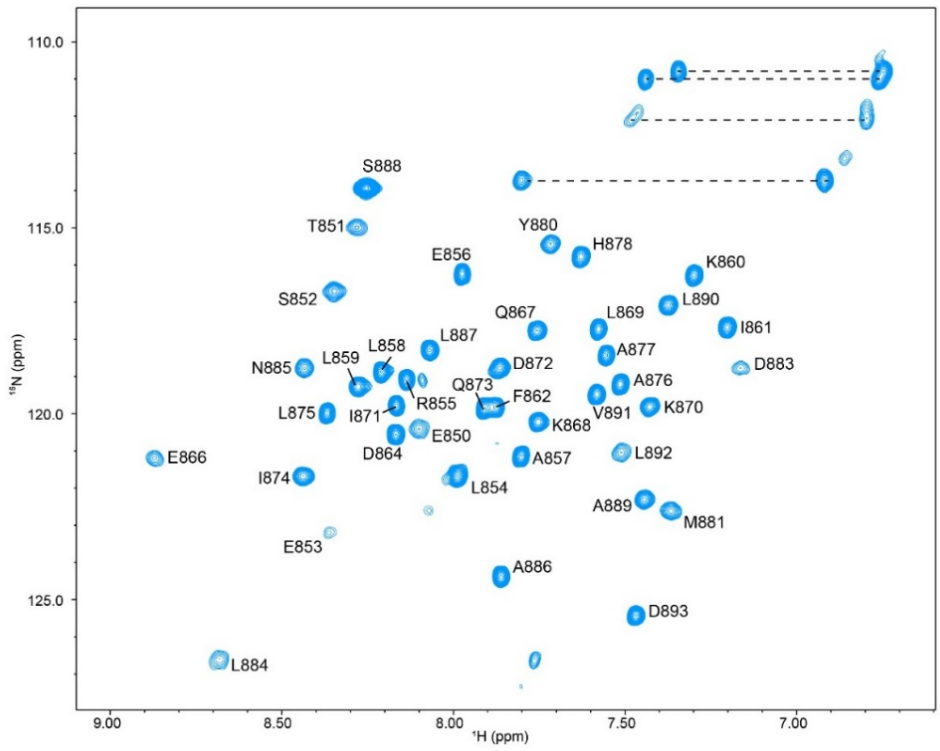


Figure 4.2.13 $^1H/^{15}N$-HSQC assignments of the N4BP1 CUE domain. Dashed lines link the sidechain amide groups of Asn or Gln which are not assigned here. x axis is the HN dimension while y axis is the NH dimension. Figure was generated by NMRviewJ.

4.2.1.2.4 Backbone Hα and side-chain H^C/C^H assignments of the CUE domain

The backbone Hα and side-chain H^C/C^H signals of the N4BP1 CUE domain were primarily collected from two 3D experiments, HBHA(CO)NH and HCCH-TOCSY. The Hα and Hβ signals of each residue were first identified in certain N^H planes of HBHA(CO)NH according to the backbone N^H and H^N values from $^1H/^{15}N$-HSQC. Furthermore, HCCH-TOCSY helps to confirm the Hα and Hβ assignments in the Cα and Cβ planes respectively. In most cases, the signals of Hγ, Hδ etc are also visible in the same CH strip. This in turn helps locate the planes of Cγ, Cδ etc (Figure 4.2.14).

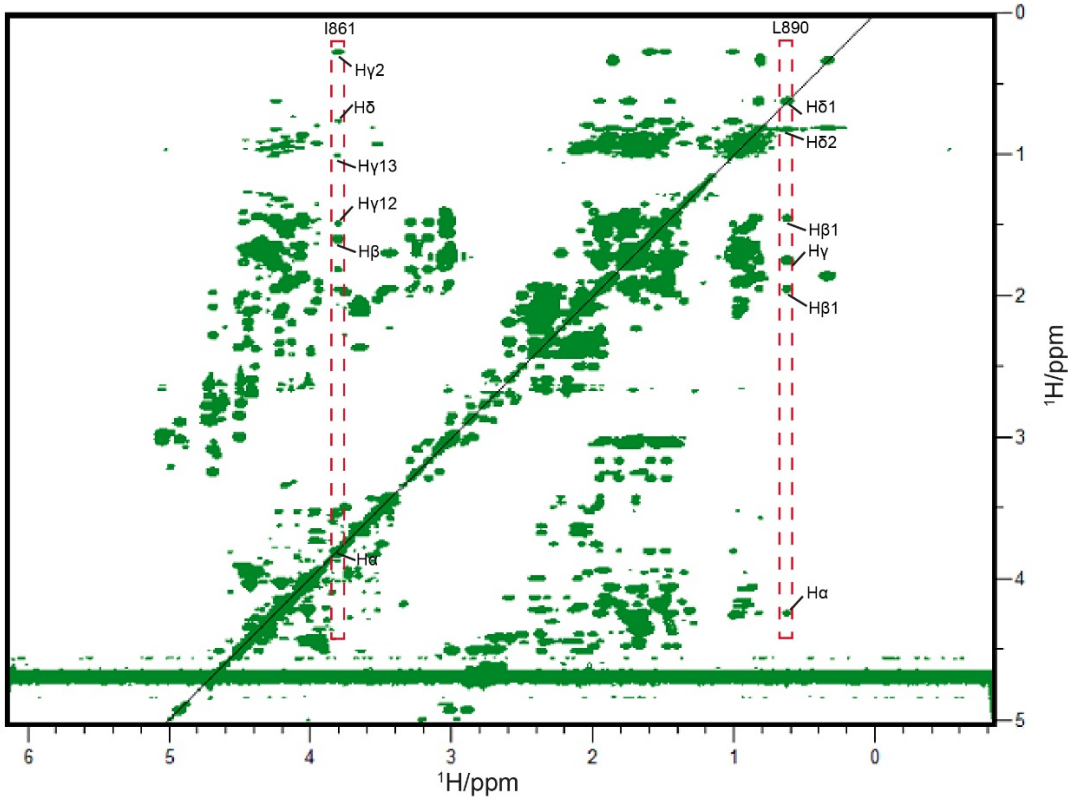


Figure 4.2.14 The compact spectrum of all ^1H-^1H planes along the C axis from HCCH-TOCSY. x axis is the direct H^N dimension while y axis is the indirect H^C dimension. The CH trips of I861 and L890 are highlighted in red dashed boxes as assignment examples. Figure was generated by CCPNMR software.

Figure 4.2.15 shows an example to demonstrate how Hα-Hδ and Cγ-Cδ were assigned for residue I861. First, the Hα and Hβ signals of I861 were identified by HBHA(CO)NH combined with $^{1}H/^{15}N$-HSQC. Then in the Cα and Cβ planes from HCCH-TOCSY, the Hα and Hβ signals are located respectively in the diagonal of the spectrum and correspond well to those in HBHA(CO)NH. The Cα and Cβ CH strip can thus be isolated. Besides, in each CH strip of either Cα or Cβ the signals of Hγ and Hδ can also be identified. Given that the direct H^N and indirect H^C values of a signal in the diagonal are the same as in HCCH-TOCSY, it is possible to find the Cγ or Cδ plane since it is known that the corresponding Hγ or Hδ must appear in the diagonal of that plane. According to this, the indirect H^C value of Hγ or Hδ was obtained from both Cα and Cβ planes, the Cγ or Cδ plane can be finally identified by going through the ^{13}C plane by matching the corresponding Hγ or Hδ signal. To conclude, HCCH-TOCSY has provided rich information on the chemical shift of sidechain H and C atoms for each residue.

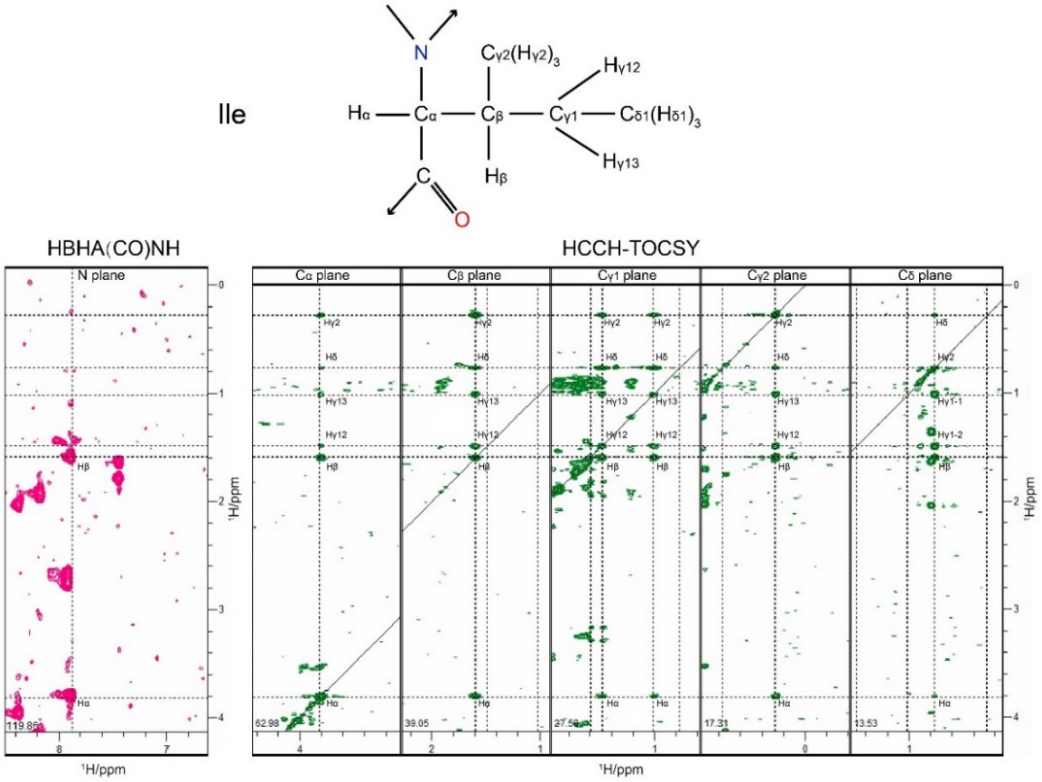


Figure 4.2.15 Backbone Hα and sidechain H^C/C^H assignments of I861 by HBHA(CO)NH and HCCH-TOCSY. Figures were generated by CCPNMR software.

In summary, almost all the backbone N^H, Hα, Cα and CO atoms have been assigned, together with most of the H^C and C^H atoms from the side chains. The overall assignment progress is indicated in Figure 4.2.16.

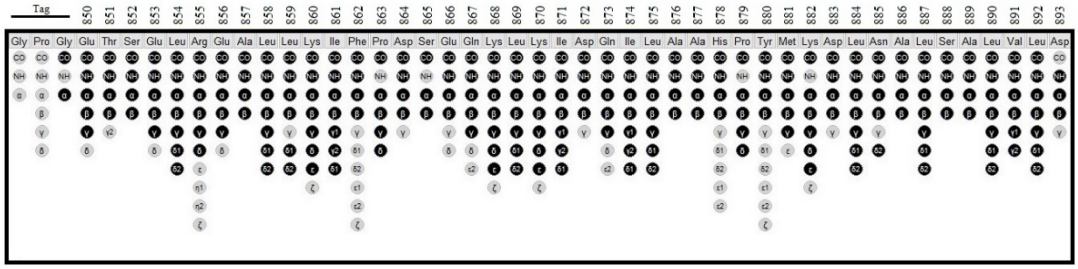


Figure 4.2.16 Overall assignment panel of the N4BP1 CUE domain. The assigned atoms are in black solid circles. The signals of the atoms in grey circle are missing in the spectra. The first 3 residues are from the tag. Figure was taken from CCPNMR.

4.2.1.2.5 Structure modelling

A full list including all the chemical shift information of available atoms from the above spectra was generated and deposited to the BMRB database (Entry: 50688). It was then submitted to CS-ROSETTA server for structure calculation. CS-ROSETTA is designed to model the *de novo* protein structure by solution NMR based on the chemical shift parameters from backbone and sidechain assignments. The modelling didn't include the proton-proton NOE constraints due to the poor quality of the ^{15}N- and ^{13}C- NOESY spectra recorded. Only a limit numbers of constraints were obtained from the NOESY spectra and they were unfortunately not good enough to reconstitute an appropriate 3-dimentional structure combined with the backbone/sidechain assignment data by ARIA. Hence, the CS-ROSETTA method was employed to establish the structure models.

The output of CS-ROSETTA is an ensemble of 10 best structural models with the lowest energy. All models are very similar to each other and converged into a single 3-dimensional structure, indicating high confidence of the model accuracy.

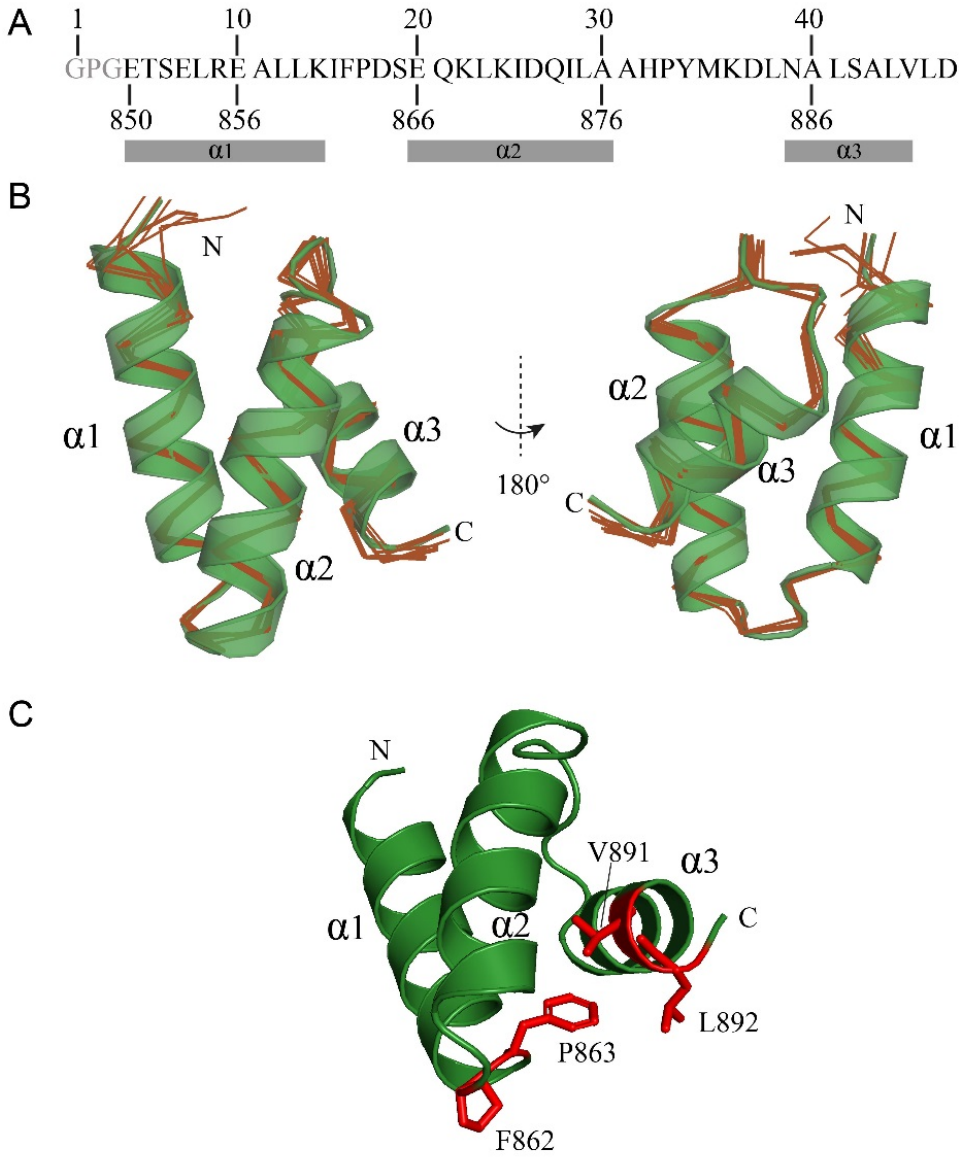


Figure 4.2.17 Best structure model of the N4BP1 CUE domain by solution NMR. A. Sequence of the CUE domain. Note that the first 3 residues "GPG" are from the tag. The numbers below the sequence indicate the positions of each residue in the full N4BP1 protein sequence. B. Structure model of the CUE domain. The best 10 structure models are indicated as brown ribbons. C. The best structure model with the lowest energy. Figures were generated by Pymol.

From the results, the N4BP1 CUE domain adopts a typical 3-helix bundle structure that is common among the classical CUE family members (Figure 4.2.17B). The first two helices have a similar length, both of which are longer

than the third helix. The weak NMR signals for the two loops indicate higher flexibility of these regions compared to the three helices. The FP motif is localized in the first loop, where the benzene ring of P863 is facing towards the third helix and in close proximity to V891 and L892. This may indicate that the key hydrophobic interaction within the molecule is important for keeping the appropriate protein structure (Figure 4.2.17C). As previously revealed from the study of the CUE domain from gp78, the FP motif is critical to keep the protein in the correct fold (Liu et al., 2012). In this study, the GST-tagged CUE F862G/P863A double mutant was expressed. Cleavage of the GST tag resulted in precipitation of the tag-free protein, indicating the unstable or disrupted structure of the F862G/P863A double mutant (data not shown).

The structure of the CUE domain will be compared with other UBDs in the discussion (section 4.2.3.1). The CUE structure allows us to characterize the ubiquitin-binding mode of the N4BP1 CUE domain by NMR titrations. In the next section, both forward and reverse NMR titrations were performed between the CUE domain and ubiquitin (including monoUb and M1/K48/K63-diUb) respectively. According to the chemical shift perturbations (CSPs) collected from each titration pair in combination with previously published ubiquitin structures, together with the CUE domain solution NMR structure solved here, an in-depth investigation of the binding surfaces was achieved in section 4.2.2.3, which further helps us understand the exact binding mode.

4.2.2.3 Quantitative analysis of the interaction between the N4BP1 CUE domain and ubiquitin probed by NMR titrations

4.2.2.3.1 Identification of the perturbed surface on the N4BP1 CUE domain upon ubiquitin binding

1) Titrating monoUb to ^{15}N-labelled CUE domain

In order to titrate the interaction by NMR, an increasing concentration of label-free monoUb was added into the ^{15}N-labelled CUE domain. The ^{1}H/^{15}N-HSQC spectra of CUE were recorded. Each signal (also called a "peak") in the ^{1}H/^{15}N-HSQC spectrum represents the backbone NH group of a certain residue (except those for side chains) from the ^{15}N-labelled CUE domain. If the peak moves upon the addition of monoUb, it indicates that the chemical environments of these residues are perturbed. These affected residues might be involved in the binding event. The amide chemical shift perturbation (CSP, usually represented by $\Delta\delta$) for each residue can be obtained when compared with the reference spectrum (the CUE domain alone, 0X spectrum in Figure 4.2.18). A list of chemical shift perturbations of all the residues can be derived from the spectra, providing rich dynamic information for multiple uses, such as calculating the equilibrium disassociation constant (Kd), mapping the binding surface, establishing the binding model by molecular dynamic simulation etc.

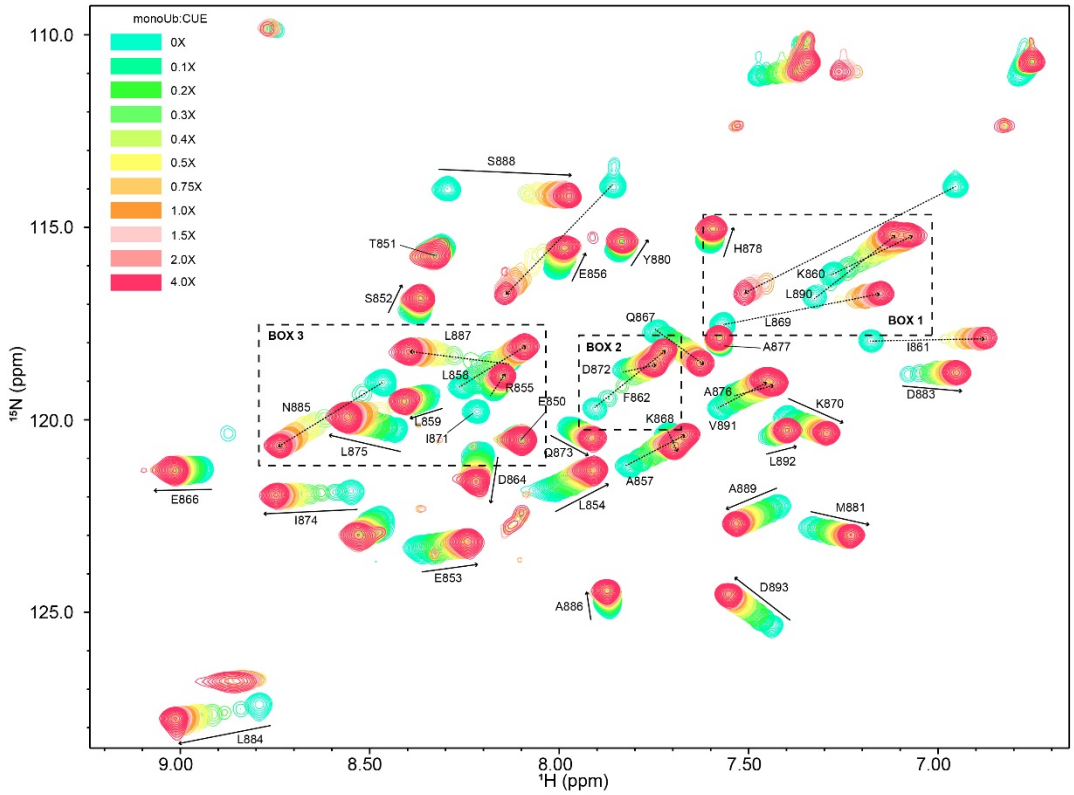


Figure 4.2.18 The serial $^{1}H/^{15}N$-HSQC spectra of ^{15}N-labelled CUE domain titrated with monoUb. The moving trend of the signal from each residue during the titration progress is indicated with an arrow. The legend indicates the molar ratio between monoUb and the CUE domain. NMR spectra were generated by NMRviewJ.

From the above titration spectra, the changes for almost all the residues can be well-tracked during the titration progress. Most of the residues were affected in the binding event, though with different degrees of chemical shift perturbations. In general, the overall signals decreased in the first several titration points (molar ratio of monoUb:CUE at 0.1X-0.5X) and then gradually went back to the original intensity. This indicates that the interaction is between fast and intermediate exchange rate. The signals started to move very slowly at 1.0X, indicating that the binding has almost reached saturation at 1:1 ratio between the CUE domain and monoUb. This confirms the binding stoichiometry from previous ITC measurements.

The chemical shift perturbation of each residue from a certain titration point was generated by the subtraction with the corresponding chemical shift value from 0X. Hence, the chemical shifts for all the residues at 0X are standardized as 0. The full chemical shift perturbation list is visualized in the chart in Figure 4.2.20A. The analysis for residue P862, S865, P879 and K882 are missing due to the absence of signals in $^{1}H/^{15}N$-HSQC. Over the titration progress, some residues have overlapped signals with each other. Therefore, a detailed peak tracking for these residues was done as shown in Figure 4.2.19. Note that the signal of I871 was completely broadened upon binding to monoUb, indicating that it undergoes strong chemical shift perturbation. In this case, the chemical shift deviation of I871 at each titration is all given at a maximum value of 0.45.

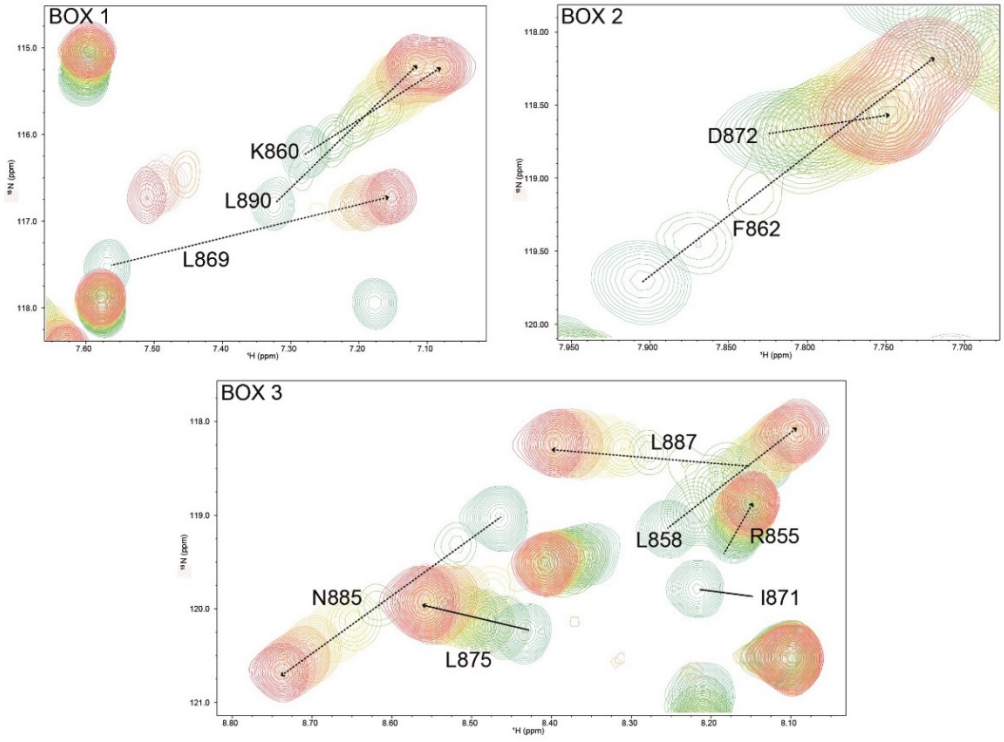


Figure 4.2.19 Enlarged local titration profiles from Figure 4.2.18. The moving trend of the serial signals from each residue is indicated by an arrow. NMR spectra generated by NMRviewJ.

With the help from AFFINImeter-NMR from the MestReNova software, the chemical shift value of each signal was isolated. The Kd for each residue can be thus calculated (the signals from side chains are not considered) (Please see more details for NMR chemical shift data fitting from section 3.2.5.4). Based on the chemical shift perturbations of all the residues, an average Kd for the CUE:monoUb titration pair was calculated with the value of 9.94 ± 1.48 μM (Table 4.2.1). It is lower than the value obtained from ITC (27.7 ± 6.73 μM). Due to limited data points for Kd calculation by NMR titration, it might not be as accurate as the ITC technique. In this case, the Kd values only derived from the NMR technique are compared in this section.

From the first sight of the CSP chart (Figure 4.2.20A), it displays a repeated "up and down" pattern among all the 3 helices, particularly in helix α1 and α3. Due to the secondary structure nature of helix, it implies that probably only a certain side of each helix is affected in the binding event. The chemical shift perturbation of each individual residue at 1.0X was mapped onto the surface of the CUE domain (Figure 4.2.20B). The key residues with strong chemical shift perturbations (CSP>0.15) highlight the possible surface areas for the binding. From an overall view, the perturbed surface is centred on the groove between helix α1 and α3. Most of the residues (L884, N885, L887, S888, L890, V891, D893) in helix α3 (including the C-terminal tail) facing to this groove are affected. The C-terminal region from helix α1 is also significantly affected upon binding and contributes to the groove. Residue F862 from the conserved FP motif in the loop between α1 and α2 is involved to build this groove. Unfortunately, residue P863 from this FP motif cannot be monitored due to that proline is invisible in $^{1}H/^{15}N$-HSQC. The last residue D893 of this domain is also moderately affected. All the above evidence is in line with the previous observation from classical CUE domains where the FP motif (from the α1-α2 loop) and a leucine residue (from helix α3) are involved in the interaction with ubiquitin.

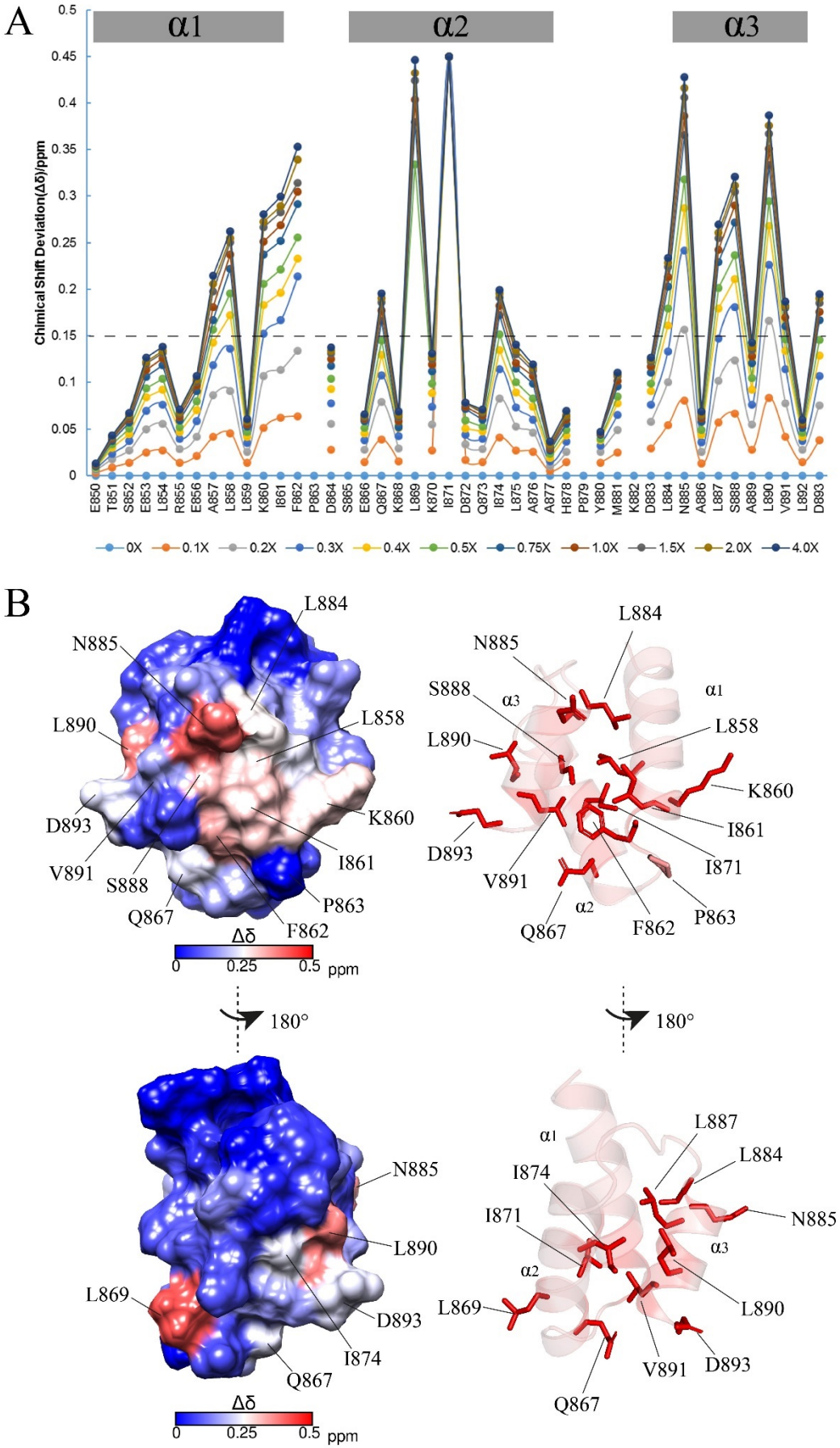

A
α1
α2
α3
Chimical Shift Deviation(Δδ)/ppm
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
E850 T851 S852 E853 L854 R855 E856 A857 L858 L859 K860 I861 F862 P863 D864 S865 E866 Q867 K868 L869 K870 I871 D872 Q873 I874 L875 A876 A877 H878 P879 Y880 M881 K882 D883 L884 N885 A886 L887 S888 A889 L890 V891 L892 D893
0X 0.1X 0.2X 0.3X 0.4X 0.5X 0.75X 1.0X 1.5X 2.0X 4.0X
B
L884
N885
L890
L858
D893
K860
V891
I861
S888
P863
Q867
F862
Δδ
0 0.25 0.5 ppm
180°
α3
α1
α2
I871
I874
L887
N885
L869

Figure 4.2.20 Chemical shift perturbation (CSP) chart from the full titration dataset (A) and the perturbed surface of CUE domain at 1.0X (B). The chemical shift deviations (Δδ) for each titration point were derived by the subtraction with the chemical shift values from 0X. The data for residue P862, S865, P879 and K882 are missing due to the lack of signals in $^{1}H/^{15}N$-HSQC. The signal of I871 was completely broadened and the chemical shift deviation at each titration is all given a value of 0.45. The chemical shift perturbation list was analyzed and generated by CCPNMR. The 1.0X chemical shift perturbation list was mapped onto the CUE domain surface by CHIMERA. The key residues with strong chemical shift perturbations (CSP>0.15) are highlighted as sticks. The cartoon views of the CUE domain were generated by Pymol.

In addition, I871 from helix α2 is pointing towards this groove (Figure 4.2.20B). The HSQC signal of I871 is completely broadened upon ubiquitin binding. It implicates that this residue also possibly takes part in the binding to ubiquitin. L869 from helix α2 also undergoes strong chemical shift perturbation though it is localized in the opposite side towards the groove. It may come from the indirect effect propagated from the perturbation of the nearby residue I871.

2) Titrating diUb chains to ^{15}N-labelled CUE domain

There are three titration pairs in these measurements, the label-free M1-diUb, K48-diUb and K63-diUb were added to the ^{15}N-labelled CUE domain at increasing concentration respectively. The changes of chemical shifts from the residues in the CUE domain upon the diUb addition were recorded by $^{1}H/^{15}N$-HSQC spectra (Figure 4.2.21A).

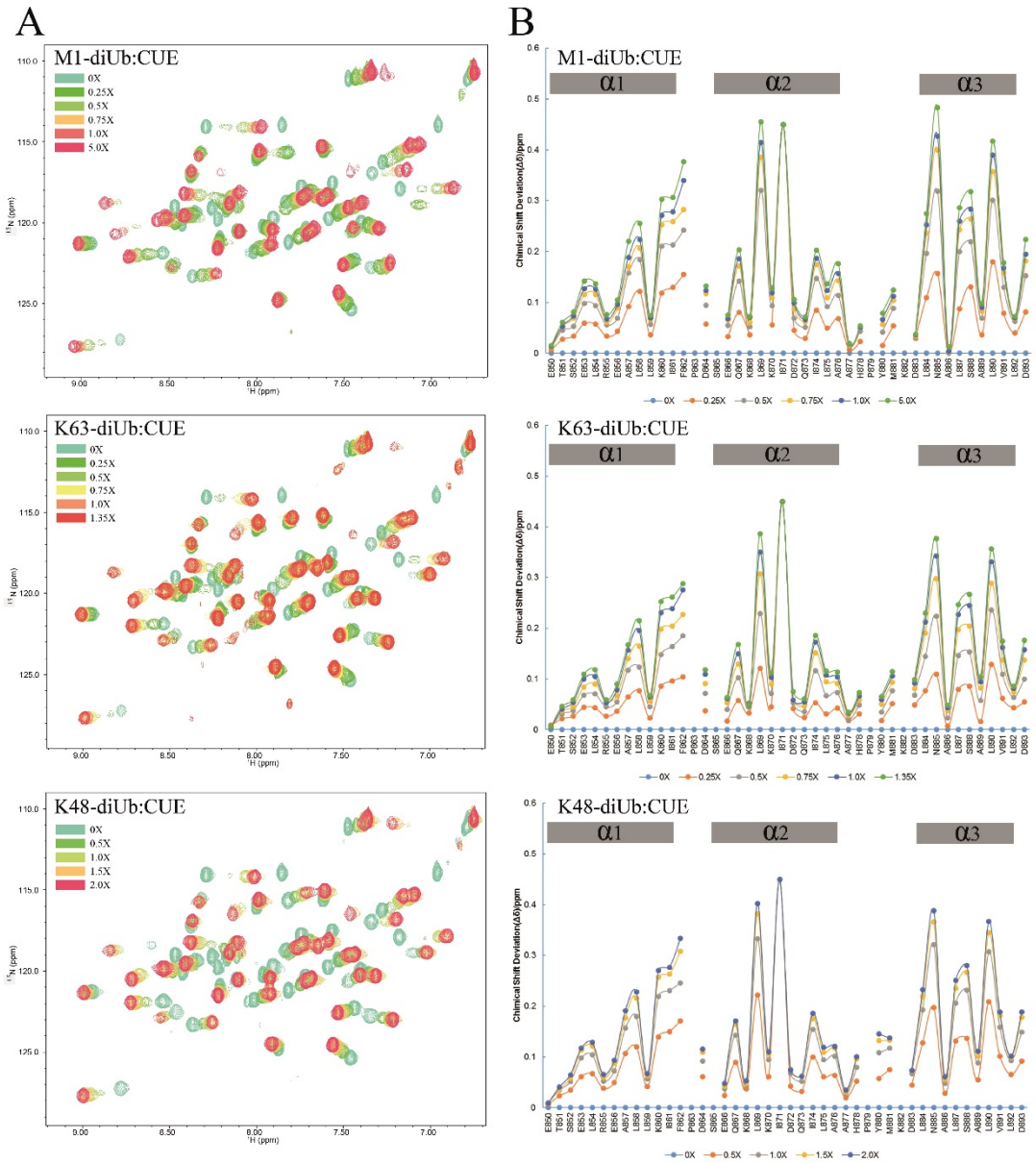


Figure 4.2.21 Chemical shift perturbations (CSPs) of ^{15}N-CUE domain upon addition of 3 different diUb chains respectively. A. Serial titration profiles by ^{1}H/^{15}N-HSQC spectra; B. The chemical shift perturbation (CSP) charts from the corresponding full titration dataset. The chemical shift perturbations (Δδ) for each titration were derived by the subtraction with the chemical shift values from 0X. The data for residue P862, S865, P879 and K882 are missing due to the lack of signals in ^{1}H/^{15}N-HSQC. The signal of I871 was completely broadened and the chemical shift deviation at each titration is all given at 0.45. The legend indicates the molar ratio between diUb and CUE domain. The chemical shift perturbation list was analyzed and generated by CCPNMR.

From the $^{1}H/^{15}N$-HSQC profiles, almost all the signals in the CUE domain upon binding to diUb are affected in the same manner within all the 3 titration pairs (Figure 4.2.21A). They are also comparable to the profile of monoUb:CUE. Quantitative analysis from the chemical shift perturbations also indicates a similar CSP pattern to each other as well as to that of monoUb:CUE, but the degrees of perturbations are different (Figure 4.2.21B, Figure 4.2.22A).

Mapping the CSPs onto the surface of the CUE domain at maximum saturation helps us visualize and compare the perturbed surface upon binding to monoUb or diUbs (Figure 4.2.22B). In general, the perturbed surfaces are almost identical. This implicates that CUE domain employs a general binding mode to interact with monoUb or diUb chains.

The CUE domain has the most affected surface when interacting with monoUb or M1-diUb. The perturbed surfaces of the CUE domain upon binding to K48-diUb and K63-diUb are slightly less affected. The average Kd values calculated for each titration pair also indicate that the CUE domain has a relatively lower affinity to K48-diUb and K63-diUb compared to monoUb and M1-diUb (Table 4.2.1).

Table 4.2.1 The averaged dissociation constant (Kd) of CUE domain upon binding to ubiquitin or chains calculated by NMR chemical shift deviations

Titration pair	monoUb	M1-diUb	K48-diUb	K63-diUb
CUE	9.94 ± 1.48 μM	10.69 ± 4.03 μM	19.22 ± 0.75 μM	17.73 ± 3.14 μM

The Kd value calculated for each individual residue was averaged to get the overall Kd. Data were fitted by the advanced tool AFFINImeter-NMR from the MestReNova software according to the method described in section 3.2.5.4.

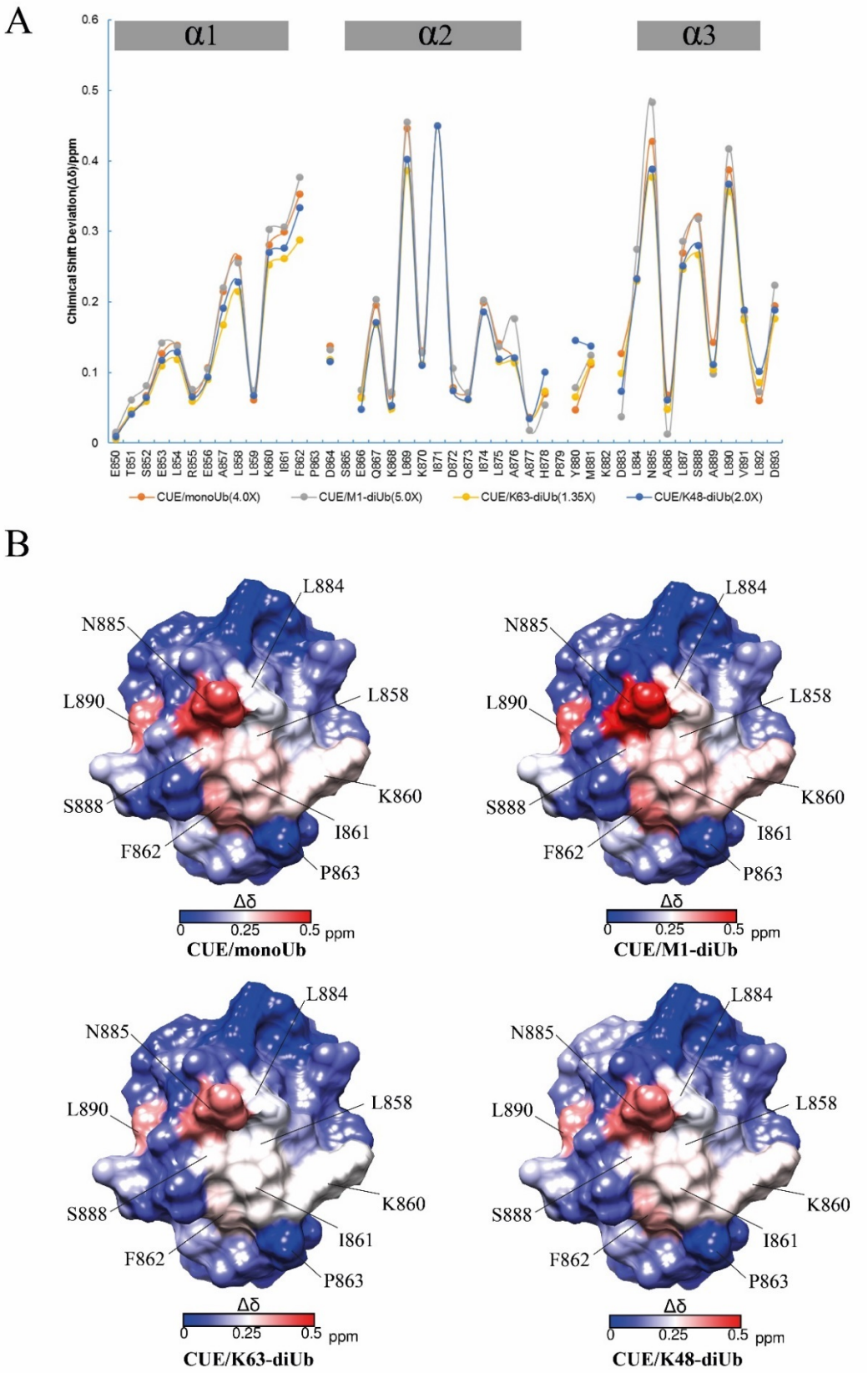


Figure 4.2.22 The chemical shift perturbation (CSP) chart for the 4 titration pairs (A) and the perturbed surface of CUE domain (B) at maximum saturation upon binding to the corresponding monoUb or diUbs. The legend in Figure A

indicates the molar ratio between monoUb (or diUbs) and the CUE domain. The maximum titration ratio for each pair is used here. The chemical shift perturbation (Δδ) for each titration were derived by subtracting the chemical shift values from 0X. The data for residue P862, S865, P879 and K882 are missing due to the lack of signals in $^{1}H/^{15}N$-HSQC. The signal of I871 was completely broadened and the chemical shift deviation at each titration is all given at 0.45. The chemical shift perturbation list at maximum saturation was analyzed and generated by CCPNMR and mapped onto the CUE domain surface by CHIMERA. The key residues with strong chemical shift perturbations are highlighted.

4.2.2.3.2 Identification of the perturbed surface on ubiquitin upon binding to the N4BP1 CUE domain

There are four titration pairs to identify the perturbed surface on the ubiquitin side. The label-free CUE domain was added to the ^{15}N-labelled monoUb, M1-diUb, K48-diUb and K63-diUb at an increasing concentration respectively. The chemical shifts of the residues from ubiquitin can be detected by recording the $^{1}H/^{15}N$-HSQC spectra. The chemical shift perturbation of ubiquitin was thus obtained by compared with the reference spectrum.

1) Titrating the CUE domain into ^{15}N-labelled monoUb

The $^{1}H/^{15}N$-HSQC profile of ^{15}N-labelled monoUb titrated with the CUE domain was the first set of measurements recorded by NMR. The full assignments of free-bound monoUb was previously introduced in Chapter II (Figure 2.6). It was used here as the reference (i.e. 0X). According to the data from ITC, the CUE domain binds to monoUb at 1:1 ratio. It is confirmed by the observation here where the peak moving became very slow or almost stopped in the $^{1}H/^{15}N$-HSQC spectra from 1.0X to 4.0X (Figure 4.2.23).

From the titration profiles, some of the peaks were moving in a certain direction to a new position during the titration process while other peaks moved very slowly or stay unchanged. In general, the intensity for most of the

signals during titration is slightly decreased but staying visible. Thus, the moving trend can be distinguished. The decreasing signal might come from the serial dilutions introduced by keeping adding the CUE domain. This suggests that the interaction between the CUE domain and monoUb is most likely undergoing a “fast exchange”, or to a less extend, between fast and intermediate exchange.

An overall Kd value of 6.44 μM was calculated for CUE:monoUb, based on those residues with significant chemical shift perturbations (as indicated in Figure 4.2.23 and Table 4.2.2).

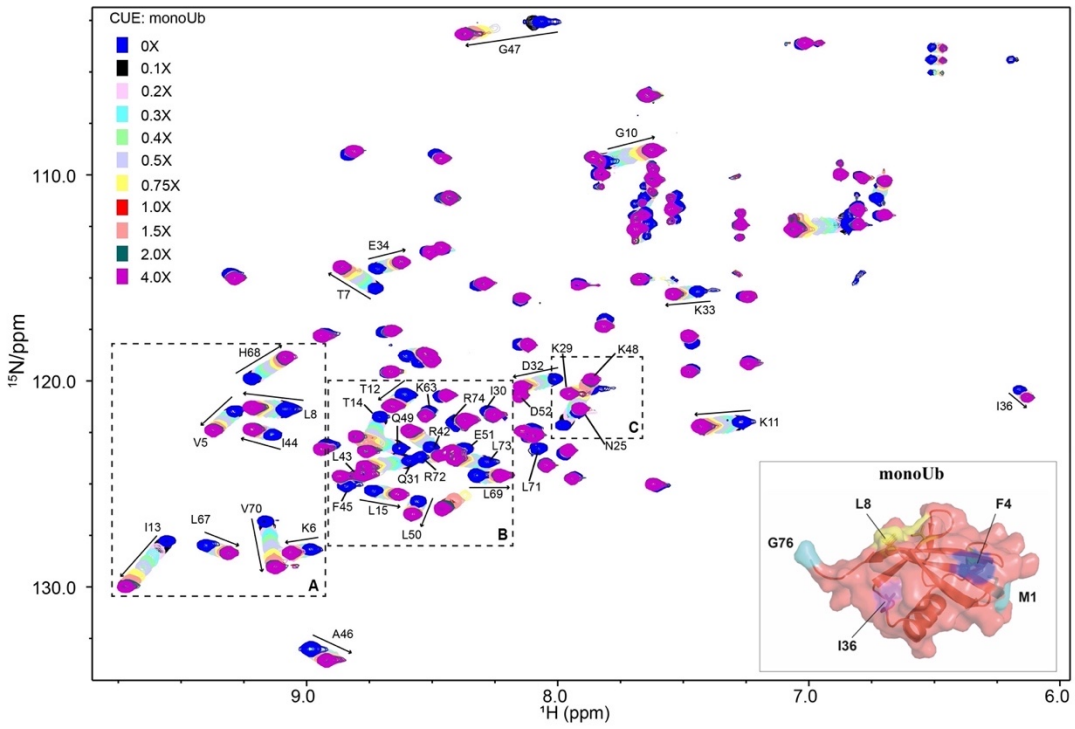


Figure 4.2.23 The serial $^{1}H/^{15}N$-HSQC spectra of 15N-labelled monoUb titrated with the CUE domain. The moving trend of the residues with significant chemical shift perturbations over the titration progress is indicated with arrows. The legend indicates the molar ratio between the CUE domain and monoUb. The structure of monoUb is inserted in the right corner. NMR spectra generated by NMRviewJ.

Residues with strong CSPs might be directly or indirectly involved in the binding event. To have a more careful analysis, some residues of interests are

selected from Figure 4.2.23 and enlarged in Figure 4.2.24 and Figure 4.2.25. In Area A, it is easy to recognize the moving tracks of all the residues as indicated by the arrows. The Kd value of each residue can be thus calculated based on their corresponding chemical shift perturbations (Table 4.2.2). There are also several residues (T14, Q31, R42, Q49, R72, L73 etc.) in Area B which need more careful efforts to distinguish (Figure 4.2.24). For example, the signals for T14 and Q49 started to merge at 0.3X and completely overlap at 0.5-0.75X. The two peaks separated again at 1X, which allows extracting the exact chemical shift pattern.

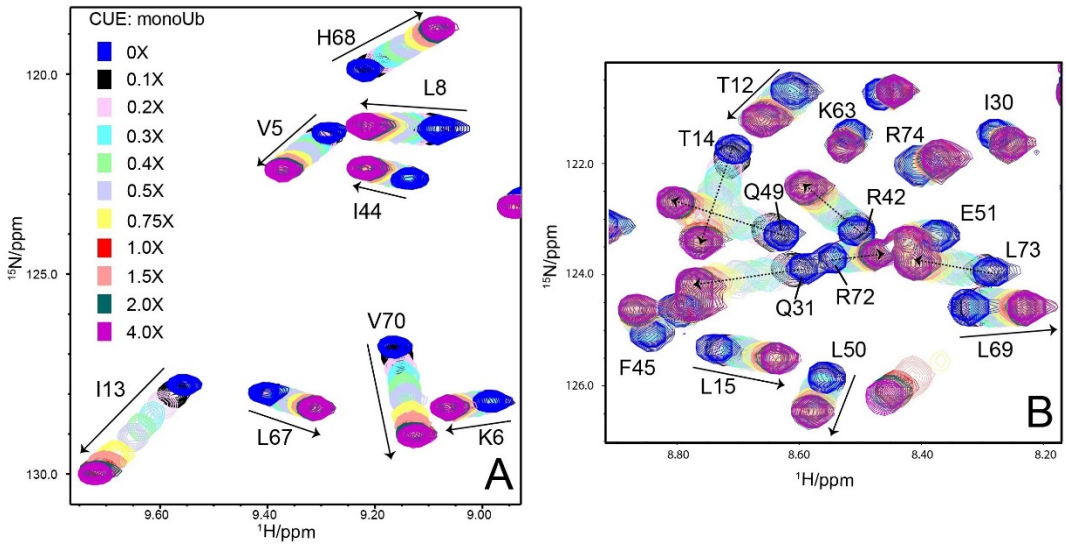


Figure 4.2.24 Enlarged local titration profiles (Area A and B) from Figure 4.2.23. The moving trend of the residues with significant chemical shift perturbations is indicated with arrows. NMR spectra generated by NMRviewJ.

It is particularly interesting to find that residue K48 was undergoing significant chemical shift perturbation during the titration (Figure 4.2.23 Area C; Figure 4.2.25). It moved across residue N25 and K29 and finally to a very distancing position, indicating that K48 is strongly affected in the binding event. The moving path of K48 is isolated while the exact peak position at 0.3-0.5X is hard to separate as indicated in Figure 4.2.25. A rough Kd calculation for K48 is ~ 6.66 μM.

Apart from the above-mentioned residues, the signal of residue L71 was completely broadened upon binding to the CUE domain, suggesting an important role in the interaction.

Table 4.2.2 Dissociation constant (Kd) of representative residues of monoUb calculated from NMR chemical shifts

Residue	Kd (μM)	Residue	Kd (μM)	Residue	Kd (μM)
V5	4.05	Q31	4.30	L50	5.93
K6	12.33	D32	5.81	E51	10.75
T7	6.03	K33	6.28	K63	7.18
L8	**5.92**	E34	5.75	L67	4.80
K11	5.65	I36	4.97	H68	3.42
T12	5.13	R42	5.85	L69	8.59
I13	5.34	**I44**	**7.53**	**V70**	**6.44**
T14	7.22	F45	3.38	R72	3.62
L15	7.50	A46	4.98	L73	4.47
N25	6.57	G47	5.62	R74	31.24
K29	8.62	**K48**	**6.66**	Average	6.44 ± 0.51
I30	7.66	Q49	5.53		

Data were fitted by the advanced tool AFFINImeter-NMR from the MestReNova software according to the method described in section 3.2.5.4. Due to large errors in the Kd calculation for each residue, an average Kd was calculated based on all the residues listed in the table to represent the overall affinity.

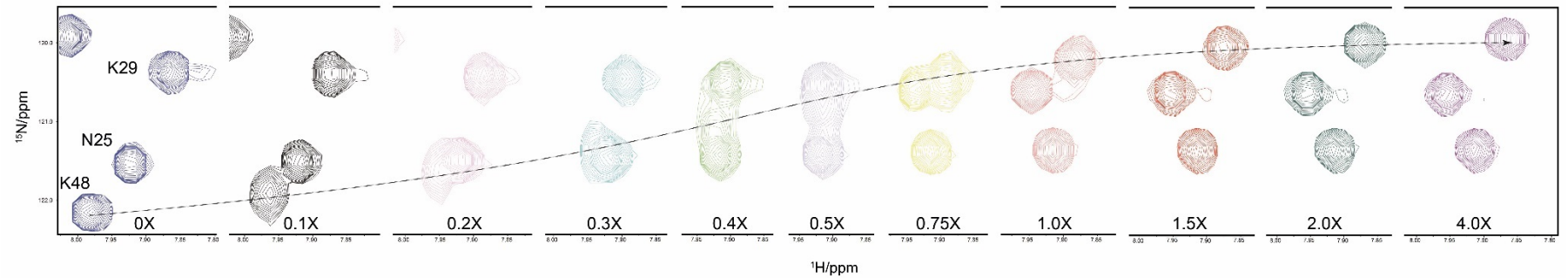


Figure 4.2.25 Resolving the signals of residue K48 during the titration progress. A dashed trendline is indicated to show the moving trend. NMR spectra generated by NMRviewJ.

2) Titrating the CUE domain into ^{15}N-labelled M1-diUb

M1-diUb consists of two mono-ubiquitin molecules linked via the normal peptide bond in a "head to tail" manner, thus this ubiquitin linkage is "linear". Due to the overall similar chemical environment of the two ubiquitin moieties in one M1-diUb molecule, it is no surprise that most of the ^{1}H/^{15}N-HSQC spectrum of free-bound M1-diUb is similar to that of monoUb (Figure 4.2.26; See Figure A1 from Appendix II for the full assignments of M1-diUb). Because of this, on the other hand, it is difficult to separate the signals of the equivalent residues from both moieties. Therefore, in most cases, the overlapped peaks in the HSQC spectrum of M1-diUb are annotated with both pairwise residues, for example, V5/81, K48/124 etc.

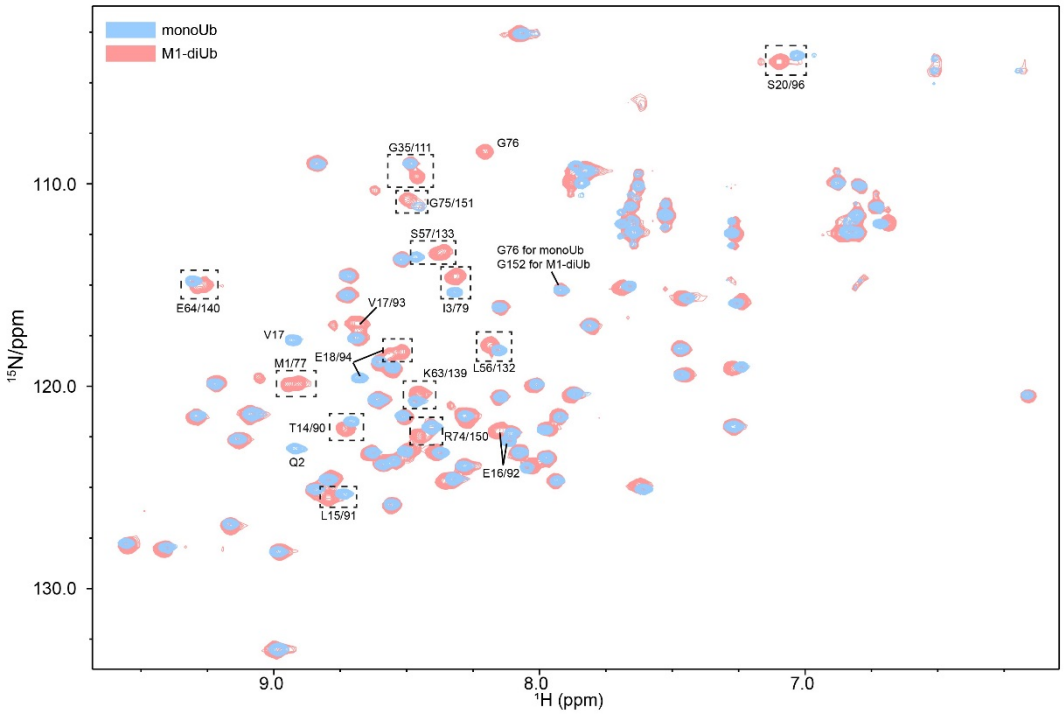


Figure 4.2.26 The overlapped ^{1}H/^{15}N-HSQC spectra of monoUb and M1-diUb at free-bound states (oX). The signals are color-coded according to the corresponding spectrum. Residues with difference chemical shifts between the two spectra are highlighted. NMR spectra generated by NMRviewJ.

Even so, there are still some differences between monoUb and M1-diUb. Some of the residues have deviated chemical shift values between the two spectra. Some of the peaks in M1-diUb split into two signals including E18/94, G35/111, R74/150, G75/151 etc, indicating the slightly different chemical environments between the two equivalent residues. Noteworthy, G76 and G152 can be well separated in M1-diUb, among which the G152 signal is overlapped with G76 from monoUb. The signal of Q2 is visible in monoUb but absent in M1-diUb, while M1/77 are visible in M1-diUb but absent in monoUb (Figure 4.2.26).

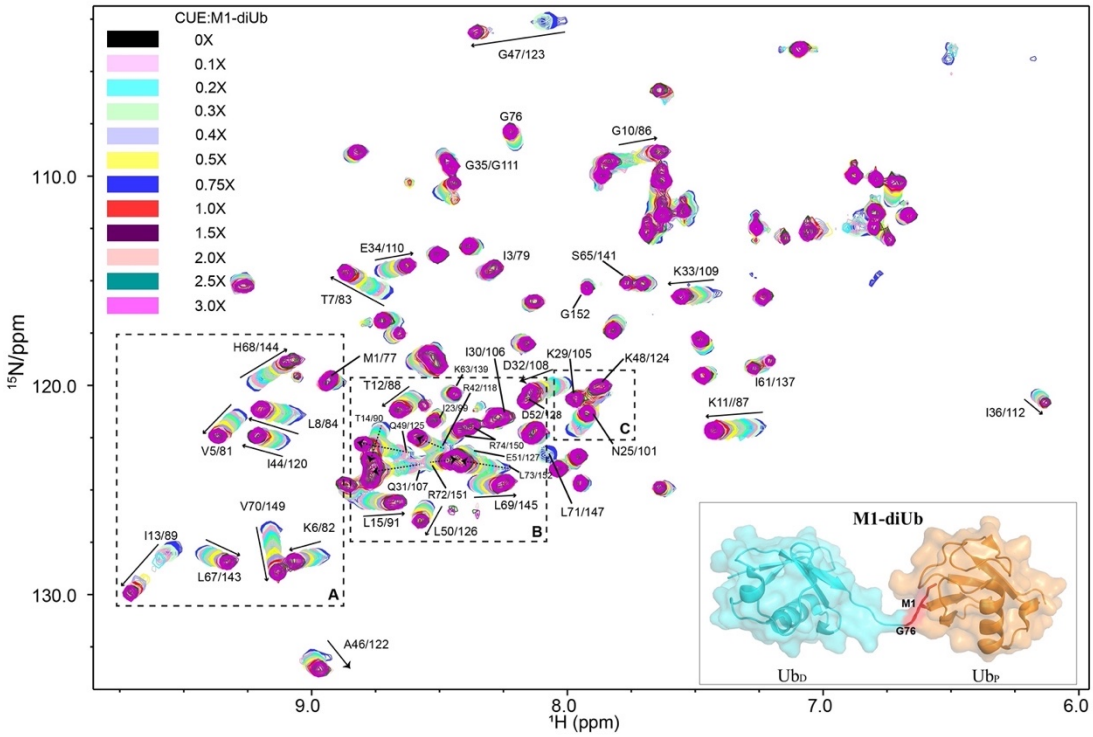


Figure 4.2.27 The $^{1}H/^{15}N$-HSQC spectra of M1-diUb titrated with the CUE domain. The moving trend of the residues with significant chemical shift perturbations during the titration progress is indicated with arrows. The legend indicates the molar ratio between the CUE domain and M1-diUb. The structure of M1-diUb is inserted in the right corner. Ub_P: proximal Ub; Ub_D: distal Ub. NMR spectra generated by NMRviewJ.

The $^{1}H/^{15}N$-HSQC spectral profile of titrating CUE into ^{15}N-labelled M1-diUb is also very similar to that of CUE into monoUb (Figure 4.2.27).

Most of the peaks moved in the same way as in CUE:monoUb and stopped moving from 2X. The spectra at 2.5X and 3X are almost identical to 2X. This observation is in line with the 2:1 binding ratio of CUE:M1-diUb from the ITC measurements.

By addition of an increasing amount of the CUE domain into ^{15}N-labelled M1-diUb, some of the residues were undergoing significant chemical shift perturbations. The Kd values of these residues are summarized in Table 4.2.3. The average Kd based on these residues is 14.77 μM to reflect the overall binding affinity between the CUE domain and M1-diUb.

Table 4.2.3 Dissociation constant (Kd) of representative residues of M1-diUb calculated from NMR chemical shifts

Residue	Kd (μM)	Residue	Kd (μM)	Residue	Kd (μM)
M1/77	9.99	Q31/107	11.48	I61/137	18.95
V5/81	9.70	D32/108	23.16	S65/141	23.61
K6/82	11.81	K33/109	15.64	L67/143	8.85
T7/83	24.12	E34/110	18.50	H68 or H144*	5.76
L8/84	**13.20**	I36/112	12.09	H144 or H68*	15.05
G10/86	12.95	R42/118	19.47	L69 or L145*	4.83
K11/87	14.13	**I44/120**	**11.53**	L145 or L69*	13.91
T12/88	13.31	A46/122	21.25	**V70/146**	**13.89**
I13/89	19.70	G47/123	39.06	R72/148	27.79
T14/90	24.32	**K48/124**	**5.16**	L73/149	33.05
L15/91	20.60	Q49/125	5.12	R74 or R150*	14.20
N25/101	30.36	L50/126	20.78	R150 or R74*	64.63
K29/105	20.73	E51/127	34.96	Average	14.77 ± 0.27

Data were fitted by the advanced tool AFFINImeter-NMR from the MestReNova software according to the method described in section 3.2.5.4. *Note that the Kd of H68 and H144, L69 and L145, R74 and R150 can be all resolved but cannot be distinguished with each other in each pair, thus cannot be assigned to a certain residue. Due to large errors in the Kd calculation for each residue, an average Kd was calculated based on all the residues listed in the table to represent the overall affinity.

Several local areas were selected from the titration profiles (Area A-C and some discrete peaks from Figure 4.2.27) for a careful investigation.

As enlarged in Figure 4.2.28, nine residue pairs are displayed in Area A. Signals from H68 and H144 were overlapped when the CUE domain was absent. However, when it came to the full saturation state (at 3.0X), the two residues can be very well separated. This indicates that the two equivalent histidine residues from the two ubiquitin moieties are affected in different degrees upon binding to the CUE domain. This “one peak to two peaks” pattern also happens to I3/79, I61/137 and S65/141.

There are also examples where both equivalent residues from the two ubiquitin moieties can be resolved. In the case of M1 and M77, the two residues can be separated from each other no matter the CUE domain was present or not, indicating the difference chemical environments around them. This can be explained by the fact that the M77 from the proximal ubiquitin has been occupied to form the linear linkage with G76 from the distal ubiquitin while M1 is free and exposed to the solvent in M1-diUb. Throughout titrations, it is observed that only one of the signals from M1 and M77 was slightly moving while another one stayed almost still, indicating that only one methionine residue in this pair is affected in the binding event.

For the rest of the residue pairs in Area A, the two equivalent residues in each pair were completely overlapped and followed a certain trend together during the titrations but cannot be separated from each other at all time. It indicates that the two equivalent residues from the two ubiquitin moieties are affected in the same fashion upon binding to the CUE domain.

In Area B, each residue in pair L69/145 and R74/150 was well separated from another during the titration progress respectively. However, they were perturbed in different degrees as reflected from the Kd values. Residue pair L15/91 and I30/106 tend to diverge from each other respectively throughout titrations, but it is hard to separate the two signals completely apart.

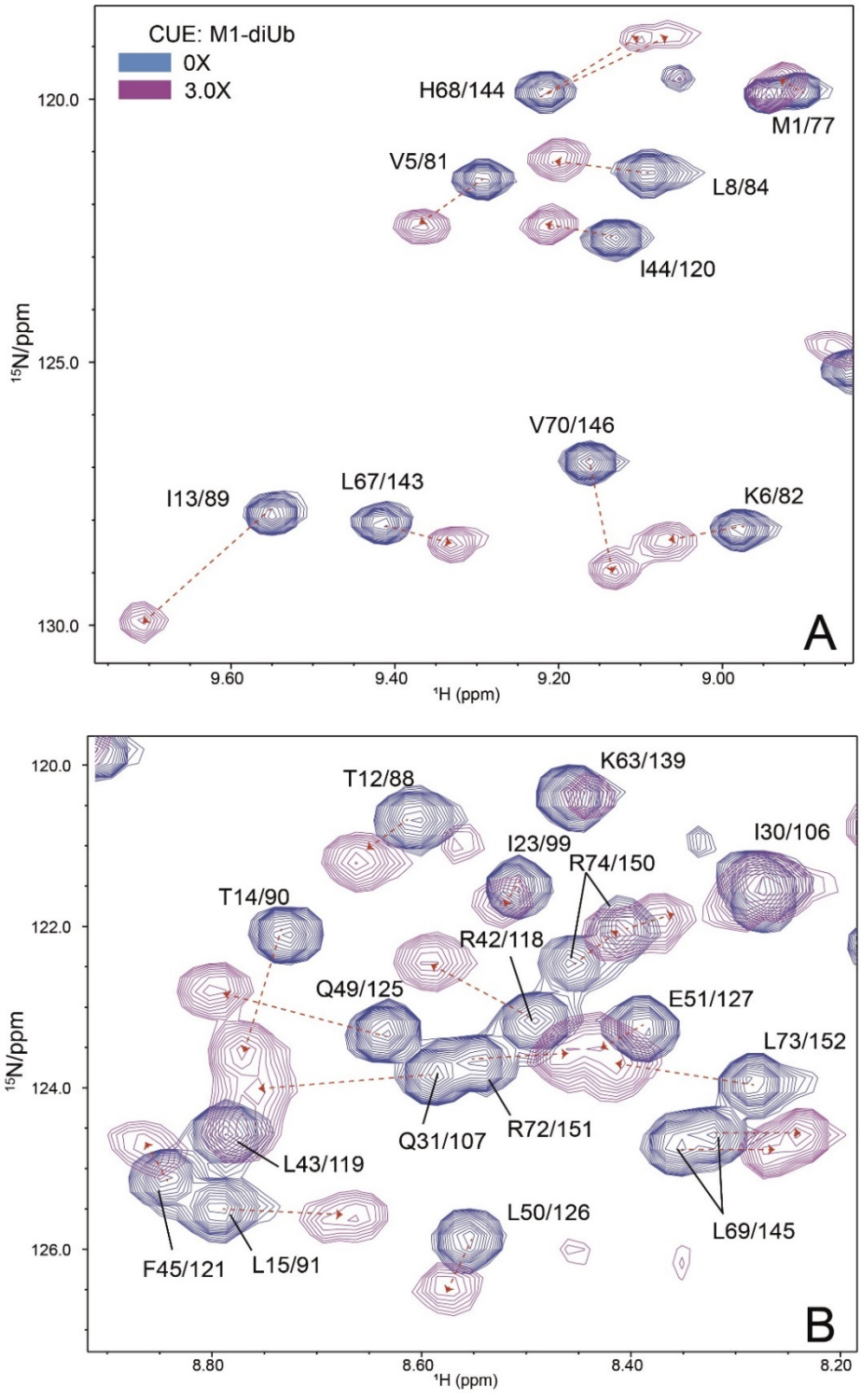


Figure 4.2.28 Enlarged local titration profiles (Area A and B) from Figure 4.2.27. The moving trends of residues from 0X to 3X with significant chemical shift perturbations are indicated with red dashed arrows. NMR spectra generated by NMRviewJ.

In Area C, three residues N25/101, K29/105 and K48/124 are very near to each other (Figure 4.2.29A). The chemical shift perturbation pattern is highly similar to that of CUE:monoUb. The signals of the equivalent residues from each pair remain overlapped throughout the titration progress. K48/124 again was undergoing strong chemical shift perturbation. While the CUE domain was increased, the signal of K48/124 evolved fast and crossed the N25/101 signal from 0.1X to 0.5X and then passed the K29/105 signal from 0.5X to 1.0X. At 1.0X, three individual peaks reappeared again, displaying a clear moving trend. K48/124 seems to start splitting into two signals from 1.0X, but unfortunately the resolution is not enough to separate them apart. From a rough calculation, the Kd for K48/124 (5.16 μM) from CUE:M1-diUb is comparable to K48 from CUE:monoUb (6.66 μM). This result further supports the importance of residue K48 from ubiquitin for the interaction with the CUE domain.

There are two residue pairs I61/137 and S65/141, following the "one peaks to two peaks" pattern as mentioned before but behave a little differently (Figure 4.2.29B). When CUE:M1-diUb reached to 1:1 ratio (i.e. 1X), it is clear to see two separated peaks from the original one. However, only one of the peaks kept moving while another kept staying still. This indicates that one of the equivalent residues from the two ubiquitin moieties was undergoing remarkable chemical shift perturbation while another was hardly affected in the binding event.

Despite the differences between the two ubiquitin moieties from M1-diUb described above, the chemical shift perturbation pattern for each ubiquitin moiety is still highly similar to that of monoUb titrated with the CUE domain (It will be compared in Figure 4.2.39). A detailed CSP comparison will also be discussed in part 5 in this section.

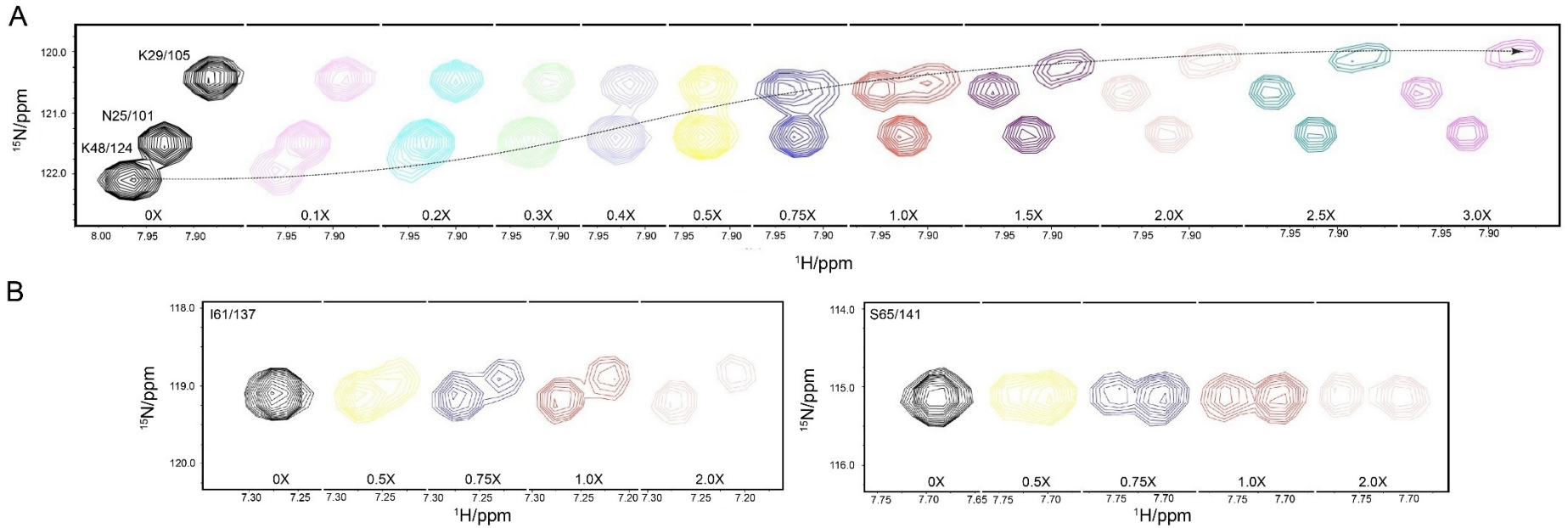


Figure 4.2.29 Resolving the signals of K48/124 (A), I61/137 and S65/141 (B) during the titration progress. Figure A is the enlarged image of Area C from Figure 4.3.16. A dashed trendline is indicated to show the moving trend. NMR spectra generated by NMRviewJ.

3) Titrating the CUE domain into ^{15}N-labelled K48-diUb

Due to that the two ubiquitin moieties are linked via K48 linkage in K48-diUb, each ubiquitin is considered as an individual chain from the structural view. Thus, the same numbering rule from monoUb (From M1 to G76) was used for each chain in K48-diUb. Unlike the similar chemical environment for both ubiquitin moieties in M1-diUb, more equivalent residues from the two ubiquitin moieties in K48-diUb can be distinguished (See Figure A2 from Appendix II for the full assignments of K48-diUb). Figure 4.2.30 shows the serial ^{1}H/^{15}N-HSQC spectral profile of titrating the CUE domain into ^{15}N-labelled K48-diUb. Although at first sight, the spectra look similar to those of CUE:monoUb or CUE:M1-diUb, however, more details for some key residues during the titrations reveal much different behaviour.

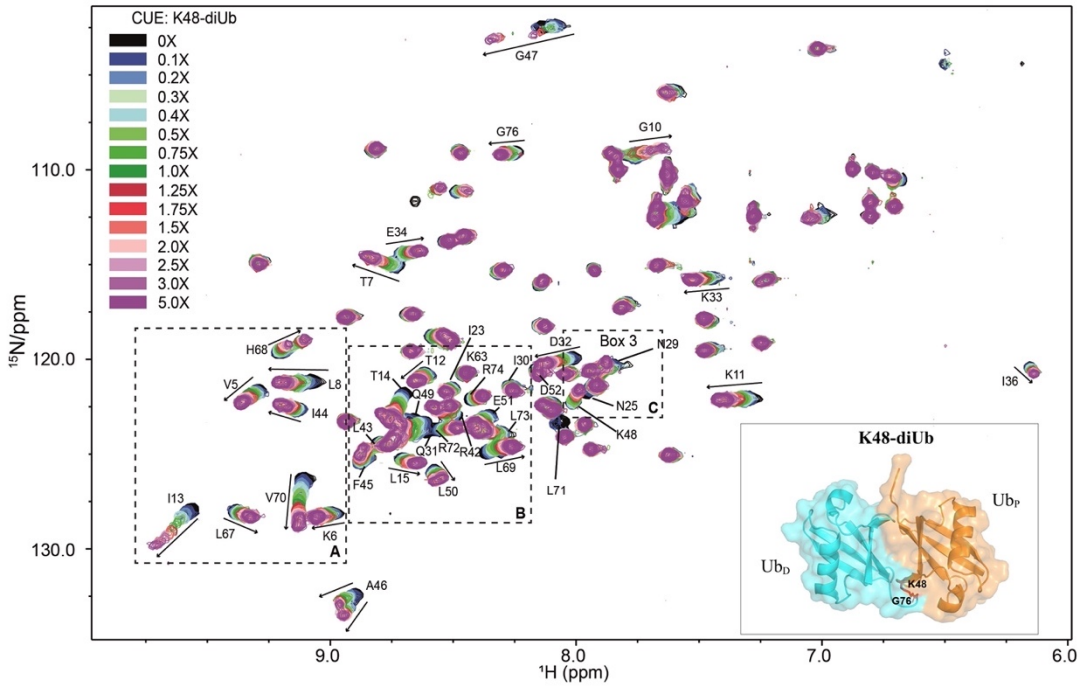


Figure 4.2.30 The serial ^{1}H/^{15}N-HSQC spectra of K48-diUb titrated with the CUE domain. The moving trends of the residues with significant chemical shift perturbations during the titration progress are indicated with arrows. The legend indicates the ratio between the CUE domain and K48-diUb. The structure of K48-diUb is inserted in the right corner. Ub$_P$: proximal Ub; Ub$_D$: distal Ub. NMR spectra generated by NMRviewJ.

To begin with, residue G47 and A46 highlight this remarkable difference.

The chemical shift changes of G47 throughout the titrations exhibited a slow to intermediate exchange pattern observed in all the titration pairs including CUE:monoUb, CUE:M1-diUb and CUE:K48-diUb. However, compared to G47/123 from M1-diUb, G47 signal from each ubiquitin chain in K48-diUb can be separated. Upon binding to the CUE domain, the signals were very weak or invisible in early stages, and then both G47 reappeared again and followed their own way to a new position respectively. Besides, one of the G47 signals moves faster than the other one. It is known that K48 in the proximal ubiquitin is occupied to form the isopeptide bond with G76 from the distal ubiquitin while K48 in the distal ubiquitin is free from forming the linkage. This different nature of the surrounding area for each G47 in K48-diUb could well explain their different perturbation degrees in the interaction. In comparison, G47 and its equivalent G123 from M1-diUb are overlapped and moved in the same way throughout the whole titration progress, similar to the behaviour of G47 in monoUb (Figure 4.2.31A).

Residue A46 is even more interesting. Apart from the well-separated signals of A46 from both ubiquitin chains in K48-diUb during the titrations, the moving direction is different from that of either monoUb or M1-diUb (Figure 4.2.31B).

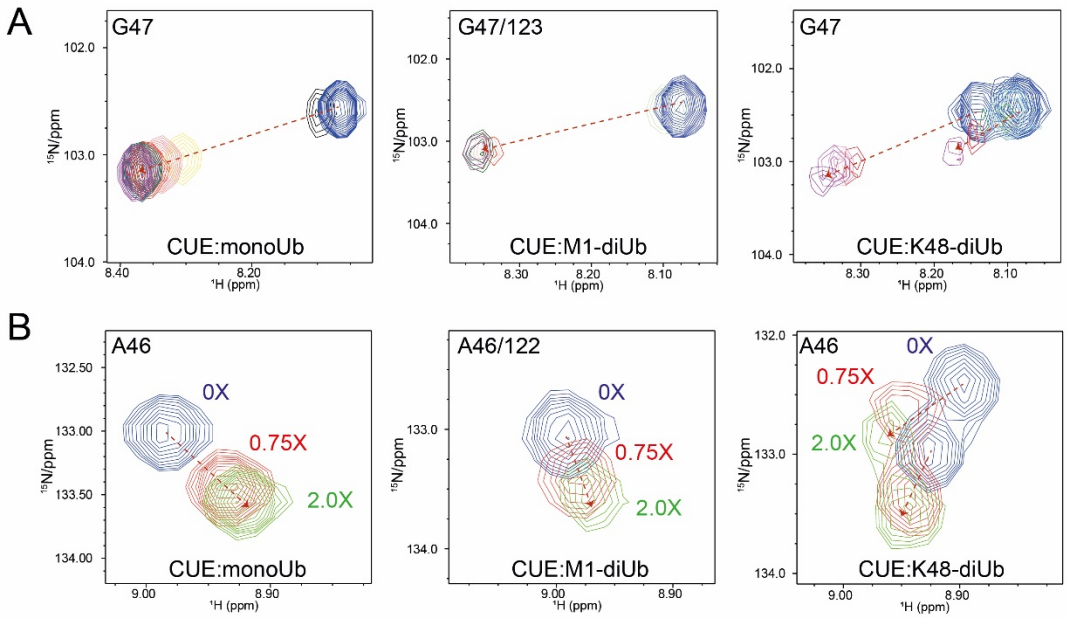


Figure 4.2.31 A comparison of chemical shift perturbation patterns for G47 (A) and A46 (B) between three different titration pairs. The moving trends of the residues over the titration progress are indicated with red dashed arrows. The signals of G47 from a full set of titration gradient (0X to maximum saturation) are shown in A. The signals of A46 at 0X, 0.75X and 2.0X are shown in B. The legend indicates the molar ratio between the CUE domain and monoUb (or M1-diUb or K48-diUb). NMR spectra generated by NMRviewJ.

Three key areas (Area A-C) from the whole spectra (Figure 4.2.30) were isolated for a more careful analysis.

Most of the residues in Area A followed the "one peak to two peaks" pattern, where the equivalent residues from the two ubiquitin chains split into two separated signals during the titrations and affected differently upon binding to the CUE domain (Figure 4.2.32).

It's worth noticing that the signals of I13 from the two ubiquitin chains can be separated very well before fully saturated by the CUE domain (for example, 0.75X as shown here). But the two signals tended to merge again at saturation (5.0X as shown here). The similar pattern also happens to residue V70 and K6, among which the two signals of K6 were completely merged at

saturation. It indicates that the CUE domain might prefer to interact with residues from a certain chain at the beginning of the binding event. Until saturation, the binding from both equivalent residues tends to reach the similar or even the same intensity.

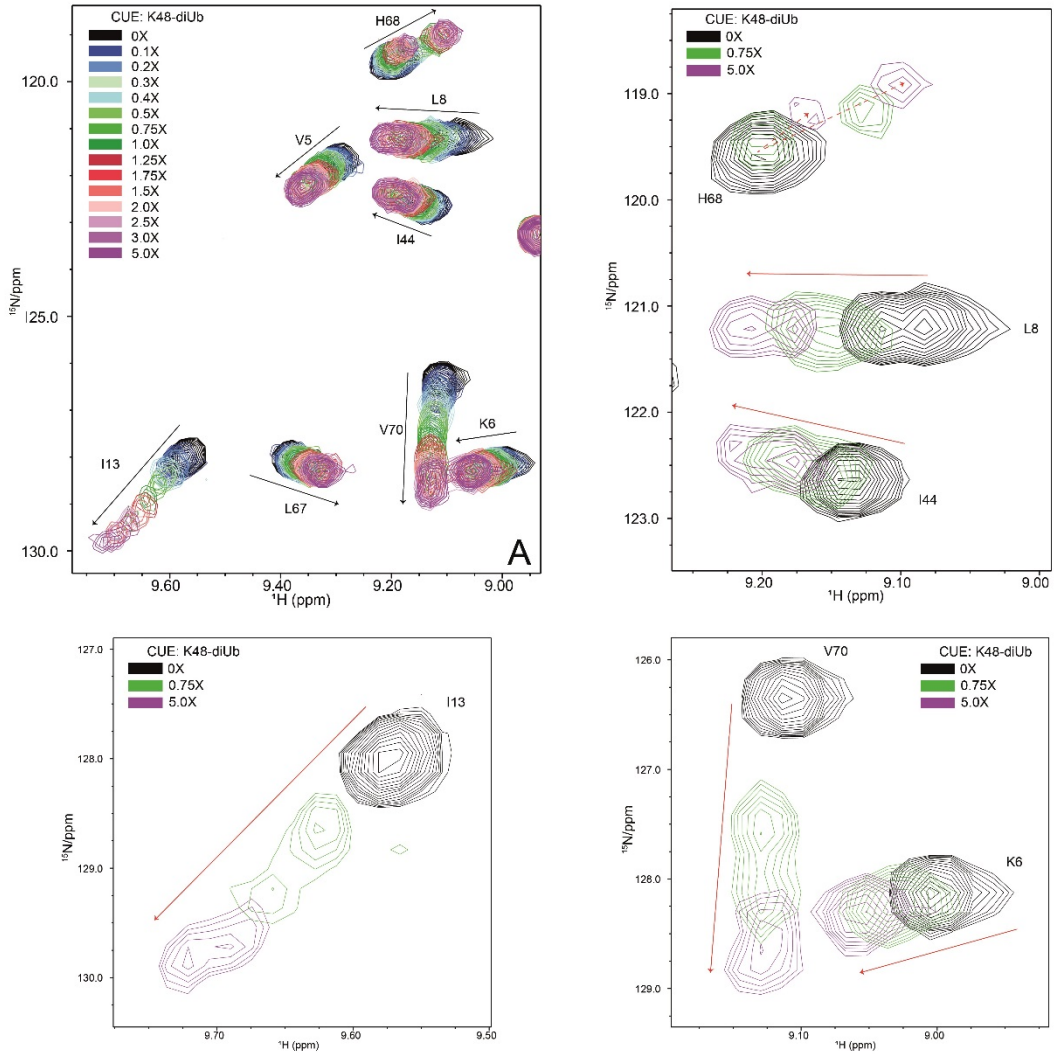


Figure 4.2.32 Enlarged local titration profile of Area A from Figure 4.2.30. Upper left corner: an overview of Area A from Figure 4.3.18. The rest of the 3 figures: the enlarged pictures of single key residues from Area A. The moving trend of each residue is indicated with black arrow or red dashed arrow. The signals are color-coded by the corresponding titration ratio as indicated in the corner. The legend indicates the molar ratio between the CUE domain and K48-diUb. NMR spectra generated by NMRviewJ.

The residues in Area B can be grouped into four classes:

Most of the residues have the "one peak to one peak" pattern, including T12, T14, I23, I30, E51, R72, L73 and R74, suggesting that the equivalent residues from both ubiquitin chains was undergoing the same perturbations via binding to the CUE domain. It's clear to see R42 and L50 followed the "one peak to two peaks" pattern, while to the opposite, the signals of Q31 and L69 from both chains started to merge respectively during the titrations ("two peaks to one peak"). F45 and Q49 displayed the "two peaks to two peaks" pattern where they had well-separated signals from both chains no matter the CUE domain was present or not (Figure 4.2.33).

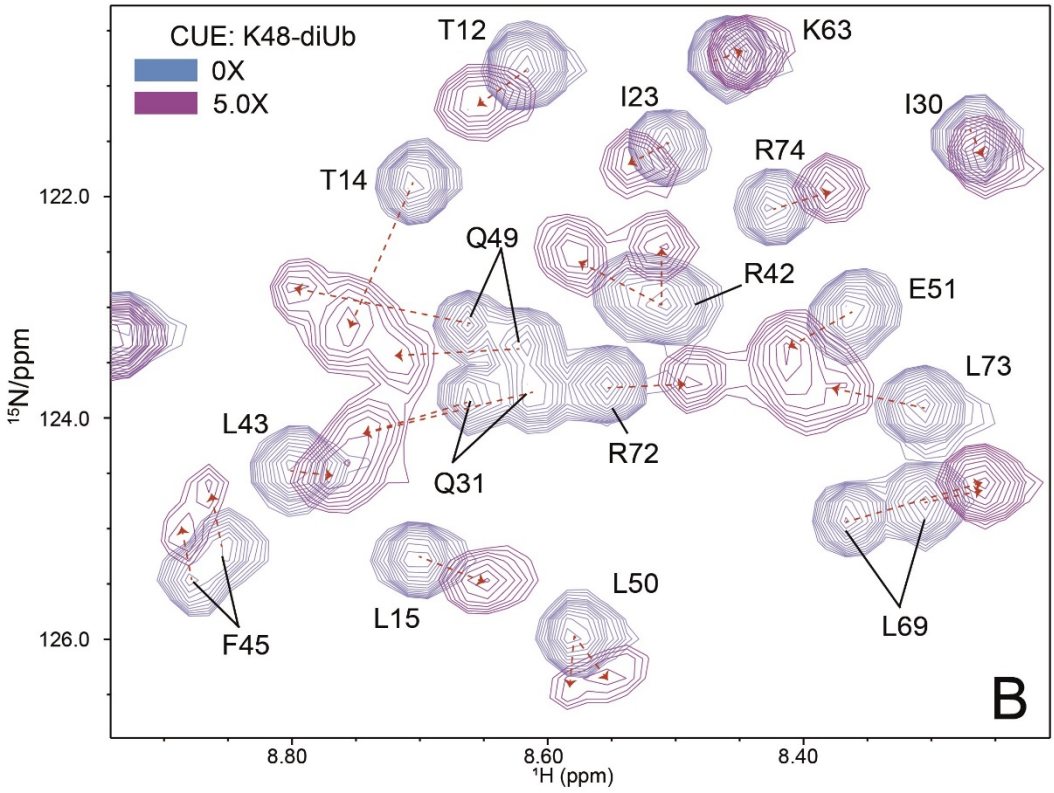


Figure 4.2.33 Enlarged local titration profile of Area B from Figure 4.2.30. The spectra from 0x and 5.0x are shown here. The moving trend of each residue is indicated with red dashed arrows. The signals are color-coded by the corresponding titration ratio as indicated in the left corner. The legend indicates the molar ratio between the CUE domain and K48-diUb. NMR spectra generated by NMRviewJ.

Residue K48 from Area C is one of the major concerns in this study. In the case of CUE:K48-diUb, the K48 signals from the two ubiquitin chains cannot be separated. It moved significantly slower than that in CUE:monoUb or CUE:M1-diUb (Figure 4.2.34). At maximum saturation (5.0X), the signal of K48 still could not pass through residue N25. This can be also well-reflected from its higher Kd value (141.1 μM) (Table 4.2.4). N25 and K29 also behaved differently from those in CUE:monoUb or CUE:M1-diUb (Figure 4.2.34). During the titration progress, the signals of N25 from both chains in K48-diUb converged to one signal while K29 diverged to two signals.

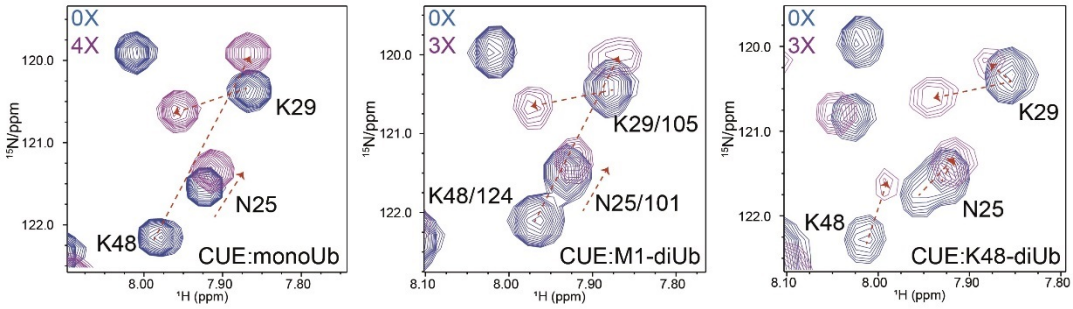


Figure 4.2.34 A comparison of residue N25, K29 and K48 among the three titration pairs. Only signals from 0X and 4X are shown here. The moving trend of each residue is indicated with red dashed arrows. The signals are color-coded by the corresponding titration ratio as indicated in the left corner of each figure. The legend indicates the ratio between the CUE domain and ubiquitin. NMR spectra generated by NMRviewJ.

The Kd values of the key residues analyzed here are summarized in Table 4.2.4. The average Kd based on these residues is 27.25 μM to reflect the overall binding affinity between the CUE domain and K48-diUb.

To summarize, there are more variations between the two ubiquitin chains in K48-diUb upon binding to the CUE domain. The average Kd for CUE:K48-diUb suggests a less tight affinity than CUE:monoUb or CUE:M1-diUb. The complete CSP list for K48-diUb also indicates a less

perturbed degree in comparison to monoUb or M1-diUb upon binding to the CUE domain (It will be compared in Figure 4.2.39).

However, from an overall view, the CSP patterns for both ubiquitin chains in K48-diUb are still largely in common with that of monoUb from monoUb:CUE (Figure 4.2.39), indicating a conserved binding mode among monoUb, M1-diUb and K48-diUb. This will be discussed in more detail together with the other titration pairs in part 5 of this section.

Table 4.2.4 Dissociation constant (Kd) of representative residues of K48-diUb calculated from NMR chemical shifts

Residue	Kd (μM)	Residue	Kd (μM)	Residue	Kd (μM)
V5	119.84	Q31-B	305.22	L50-A	41.11
K6	49.45	D32	118.30	L50-B	106.37
T7	357.94	K33	112.50	E51	13.96
L8-A	**14.84**	E34	76.33	K63	13.01
L8-B	**108.10**	I36	8.66	L67	92.66
G10	599.15	R42-A	9.59	H68-A	88.48
K11-A	165.67	R42-B	92.78	H68-B	277.40
K11-B	416.27	**I44-A**	**55.19**	L69-A	32.59
T12	67.27	**I44-B**	**74.26**	L69-B	63.15
I13-A	84.31	F45-A	8.83	**V70-A**	**18.64**
I13-B	228.64	F45-B	73.13	**V70-B**	**27.32**
T14	105.42	A46-A	17.81	R72	37.24
L15	49.82	A46-B	102.72	L73	214.30
N25-A	18.10	G47-A	138.47	R74	52.84
N25-B	29.62	G47-B	148.68	G76	47.17
K29-A	20.70	**K48**	**141.10**		
K29-B	103.40	Q49-A	84.80	Average	27.25 ± 0.47
Q31-A	79.75	Q49-B	156.73		

Data were fitted by the advanced tool AFFINImeter-NMR from the MestReNova software according to the method described in section 3.2.5.4. A and B indicate the residues from ubiquitin chain A or chain B in K48-diUb. Due to large errors in the Kd calculation for each residue, an average Kd was calculated based on all the residues listed in the table to represent the overall affinity.

4) Titrating the CUE domain into ^{15}N-labelled K63-diUb

The titrations between K63-diUb and the CUE domain are the last set of experiments. In general, the overall ^{1}H/^{15}N-HSQC spectrum of K63-diUb at free-bound state is highly similar to that of M1-diUb. This can be well-explained by the similar flexible and stretching structure of both K63-diUb and M1-diUb. The overlapped spectra suggest that the two ubiquitin moieties from both diUb molecules have very similar chemical environments. The differences between the two diUbs are limited, for example, the signals of Q2, V17 and E18 are visible in K63-diUb but absent in M1-diUb. The signal of E64/140 cannot be separated in M1-diUb, but well split in K63-diUb (Figure 4.2.35).

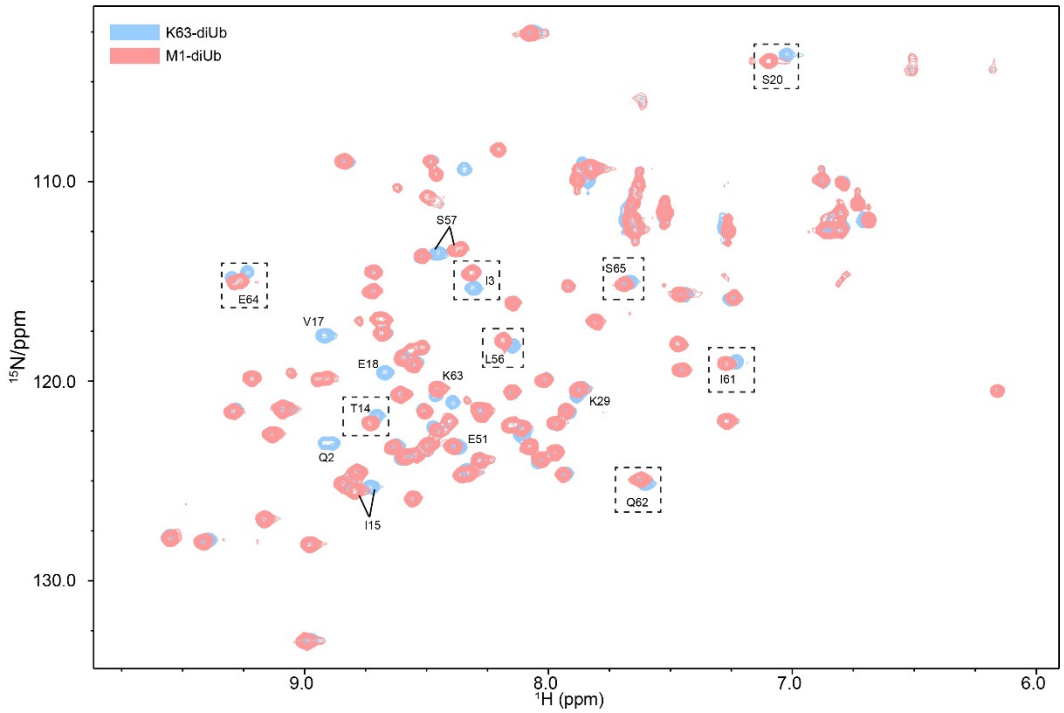


Figure 4.2.35 The overlapped ^{1}H/^{15}N-HSQC spectra of K63-diUb and M1-diUb at free-bound states (oX). The signals are color-coded according to the corresponding spectrum. Residues with difference chemical shifts between the two spectra are highlighted. NMR spectra generated by NMRviewJ.

When it comes to the NMR titrations, CUE:K63-diUb is also very similar to CUE:M1-diUb in terms of not only the moving trend but also the degrees of

chemical shift perturbations (Figure 4.2.36). The major differences between CUE:K63-diUb and CUE:M1-diUb will be discussed below.

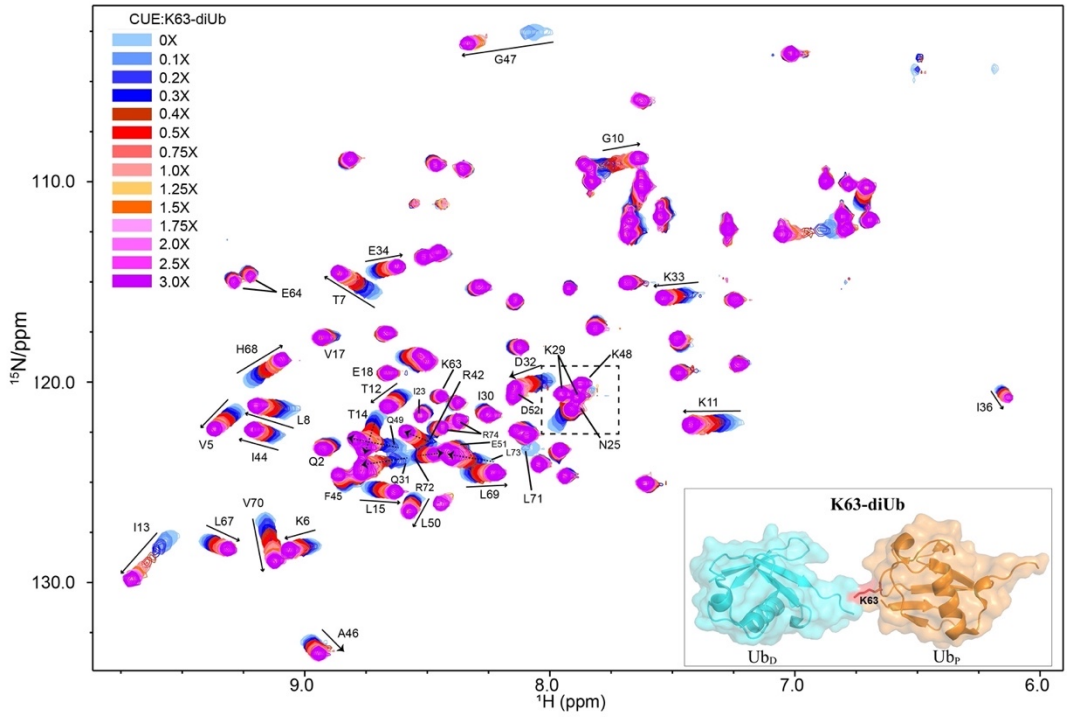


Figure 4.2.36 The serial $^{1}H/^{15}N$-HSQC spectra of K63-diUb titrated with the CUE domain. The moving trends of the residues with significant chemical shift perturbations during the titration progress are indicated with arrows. The titration ratio in the legend indicates the molar ratio between the CUE domain and K63-diUb. The structure of K63-diUb is inserted in the right corner. Ub_P: proximal Ub; Ub_D: distal Ub. NMR spectra generated by NMRviewJ.

Several key residues in K63-diUb that behave differently from M1-Ub are displayed in Figure 4.2.37. Residue Q2 is visible in K63-diUb but absent in M1-diUb. The two equivalent Q2 from both ubiquitin chains in K63-diUb can be separated at 0X, but fully converged from 0.4X until saturation (3.0X). The signal of E64/140 cannot be separated in M1-diUb while the two equivalent E64 from both chains in K63-diUb moved similarly and stayed separated at all time during the titrations. Unlike M1-diUb, two H68 signals

from both chains in K63-diUb kept being overlapped no matter the CUE domain was present or not. This happens the same to residue L69.

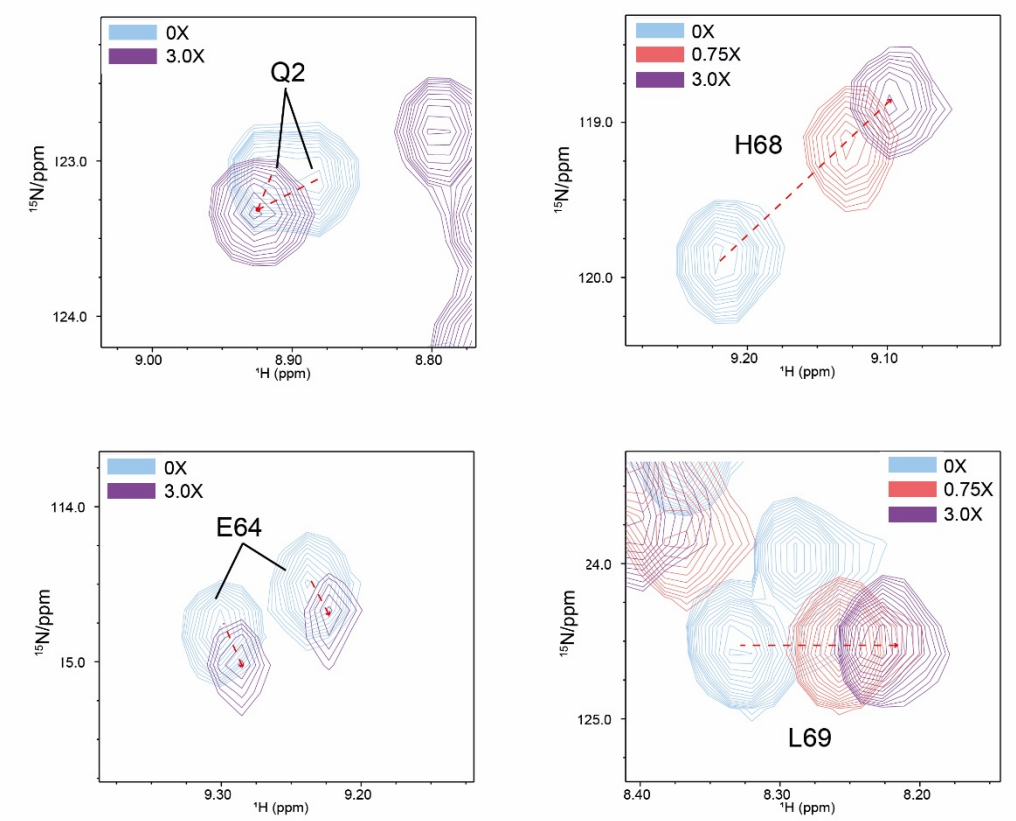


Figure 4.2.37 Resolving key residues in K63-diUb that are different from M1-diUb during the titrations. The moving trend of each signal is indicated with a red dashed arrow. The signals are color-coded by the corresponding titration ratio as indicated in the corner of each figure. The legend indicates the molar ratio between the CUE domain and K63-diUb. NMR spectra generated by NMRviewJ.

Residue K48 evolved in the same way with strong chemical shift perturbation as that in both monoUb and M1-diUb upon binding to the CUE domain. The Kd values for K48 in CUE:monoUb, CUE:M1-diUb and CUE: K68-diUb are comparable within a range of 5-8 μM, which again confirms the importance of K48 in the binding event. The signals of K29 from both ubiquitin chains in K63-diUb can be isolated at 0X. During the titration progress, K29 from one of the chains stayed unchanged while the other one was perturbed (Figure 4.2.38).

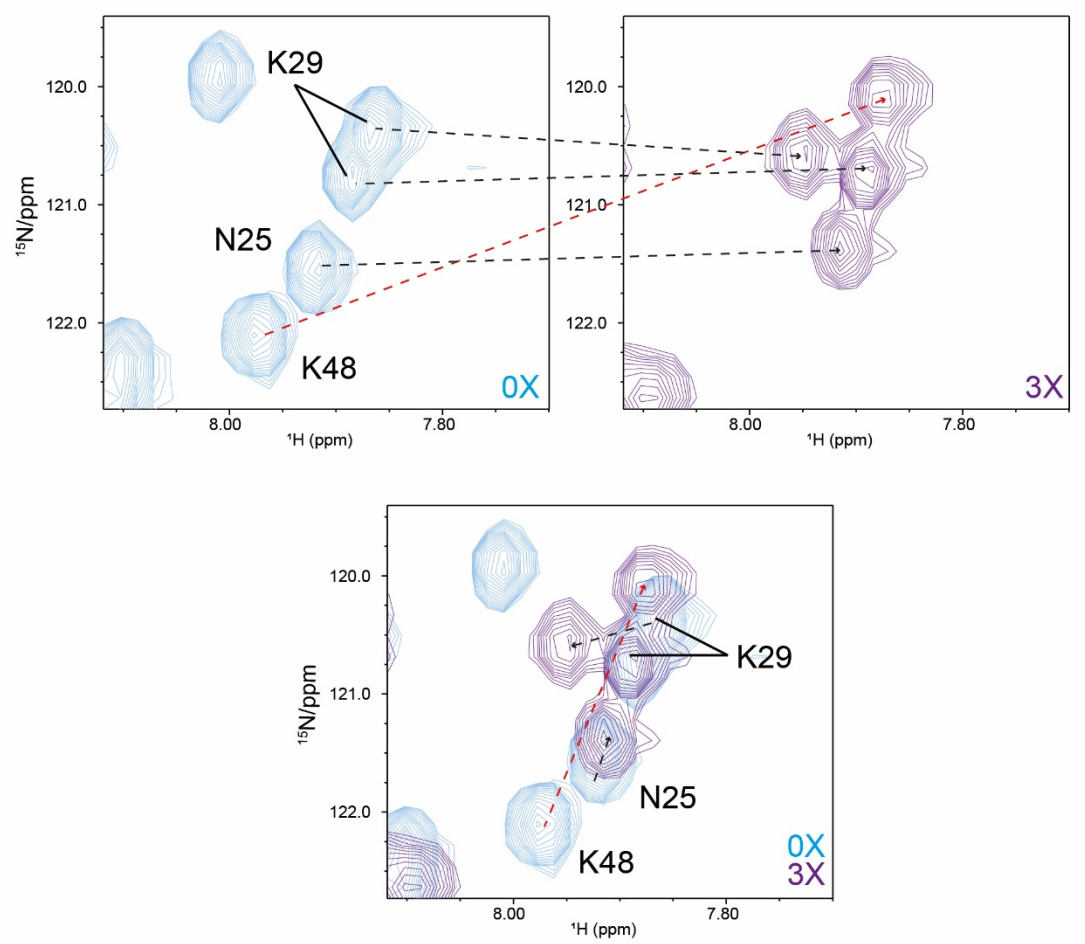


Figure 4.2.38 Resolving residue N25, K29 and K48 from K68-diUb upon binding to the CUE domain. Upper panel: the isolated spectra at 0X and 3X. Bottom panel: the overlapped spectra of both 0X and 3X. The moving trend of each signal is indicated with a red dashed arrow. The signals are color-coded by the corresponding titration ratio as indicated in the bottom right corner of each figure. The legend indicates the ratio between the CUE domain and K63-diUb. NMR spectra generated by NMRviewJ.

The Kd values of the key residues with significant chemical shift perturbations are summarized in Table 4.2.5. The average Kd based on these residues is 9.56 μM to reflect the overall binding affinity between the CUE domain and K63-diUb. This overall Kd is close to that of CUE:monoUb (6.44 μM) and CUE:M1-diUb (14.77 μM). A comparison of the CSP lists between K63-diUb and monoUb or M1-diUb also indicates the similar perturbed patterns upon binding to the CUE domain (It will be compared in Figure

4.2.39). Despite several subtle differences, it still suggests that K63-diUb employs the same binding mode for the interaction with the CUE domain.

Table 4.2.5 Dissociation constant (Kd) of representative residues of K63-diUb calculated from NMR chemical shifts

Residue	Kd (μM)	Residue	Kd (μM)	Residue	Kd (μM)
V5	8.82	K29*	15.12	**K48**	**7.97**
K6	10.34	Q31	1.942	Q49	33.85
T7	10.06	D32	11.86	L50	3.08
L8	**13.52**	K33	6.88	E51	12.21
G10	10.66	E34	12.55	L67	3.34
K11	9.47	I36	11.65	H68	14.55
T12	10.81	R42	10.24	L69	9.81
I13	7.81	**I44**	**8.88**	**V70**	**8.29**
T14	15.40	A46	10.25	R72	12.72
L15	11.38	G47	24.14	Average	9.56 ± 0.24

Data were fitted by the advanced tool AFFINImeter-NMR from the MestReNova software according to the method described in section 3.2.5.4. Note that all the residues (except K29) cannot be distinguished between chain A and B in K63-diUb. *K29 from one of the ubiquitin chains stays unchanged upon binding, hence only the one with significant CSP is shown here. Due to large errors in the Kd calculation for each residue, an average Kd was calculated based on all the residues listed in the table to represent the overall affinity.

5) Correlating chemical shift perturbations with the structure of ubiquitin

A general comparison for the perturbed surfaces of monoUb or diUb chains at saturation states from all the four titration pairs was carefully performed according to the CSPs collected from the NMR titrations.

Figure 4.2.39 summarizes the corresponding CSPs for the four titration pairs at saturation states. Residues with undetectable signals in the spectrum were left blank in the chart. Residue L71 is broadened in all the four ubiquitins. Thus, a maximum value was given. To make it easier for the comparison between monoUb and diUb chains, M1-diUb is separated into two chains. To specify, chain A is the distal ubiquitin and chain B is the proximal ubiquitin for all the three diUb chains.

From an overall view, the CSP patterns are all similar within the four ubiquitins. Most of the perturbed areas are within β2 and β4, as well as a short loop area next to the β3 C-terminus. The less affected areas include β1, α1 and α2. The above areas contain some hotspots for the binding, including the canonical I44 hydrophobic patch. Apart from the I44 patch, residue G47 and K48 which are located next to β3 also have strong CSPs in monoUb, M1-diUb and K63-diUb, but much less pronounced in K48-diUb.

Due to the overlapped signals for most of the equivalent residues in M1-diUb and K63-diUb, both proximal and distal ubiquitin moieties are affected almost in the same degree with only minor differences. The CSP intensity is also comparable to that of monoUb. However, K48-diUb has a lower level of CSPs for both ubiquitin chains compared to monoUb or M1- or K63-diUb. This is also supported by the lower affinity obtained from ITC. Due to more different chemical environments between the two ubiquitin chains in K48-diUb, the distal ubiquitin was perturbed a little stronger than the proximal one.

In brief, based on the four CSP charts, monoUb, M1-diUb and K63-diUb chains employ the same binding mode to interact with the CUE domain,

while although K48-diUb also uses a similar binding mode, there are more differences between the two ubiquitin moieties and its interaction with the CUE domain is weaker.

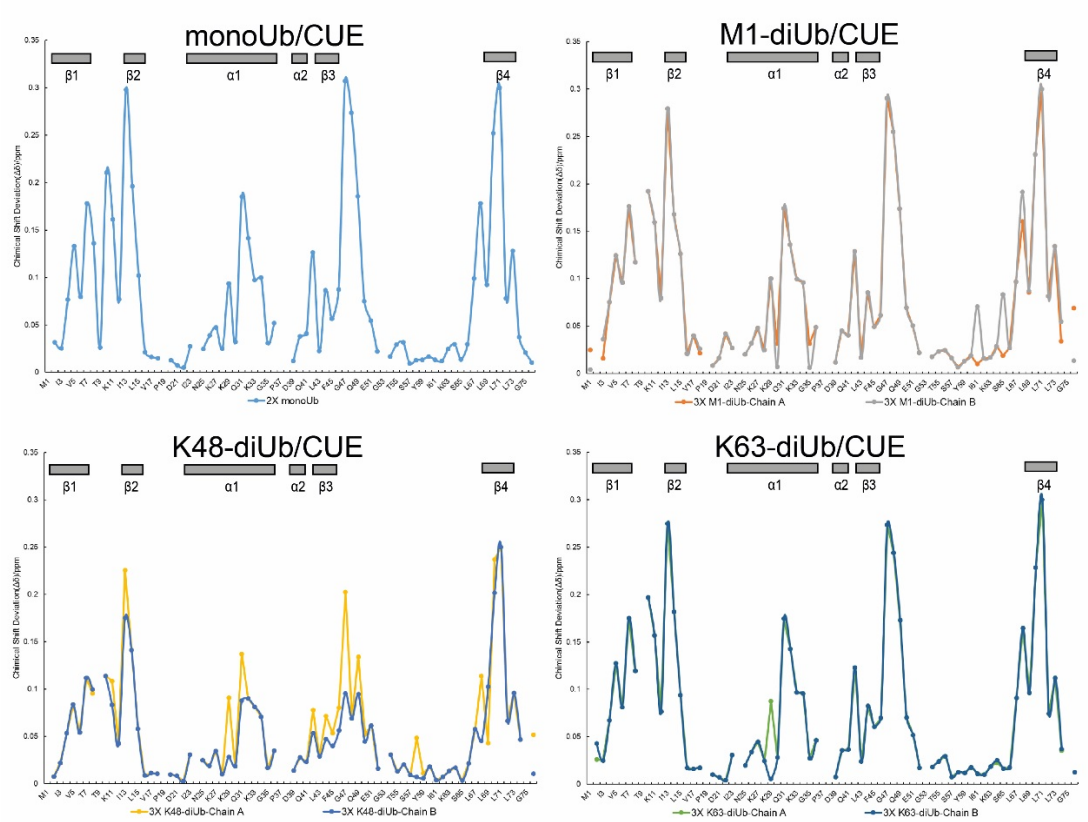


Figure 4.2.39 Chemical shift perturbations (CSPs) charts of monoUb and three diUb chains upon binding to the CUE domain at saturation states. The chemical shift perturbations (Δδ) for each titration were derived by the subtraction with the chemical shift values from 0X. Residues with missing assignments in $^{1}H/^{15}N$-HSQC are left blank. The titration ratio in the legend indicates the molar ratio between the CUE domain and ubiquitins. The chemical shift perturbation lists were analyzed and generated by CCPNMR.

The above CSPs were then mapped to the surface of monoUb and diUb chains respectively. To start with, monoUb was first analyzed (Figure 4.2.40). Residue G10, I13, G47, K48, V70 and L71 all have a CSP value above 0.2, which highlight the perturbed surface into two main areas.

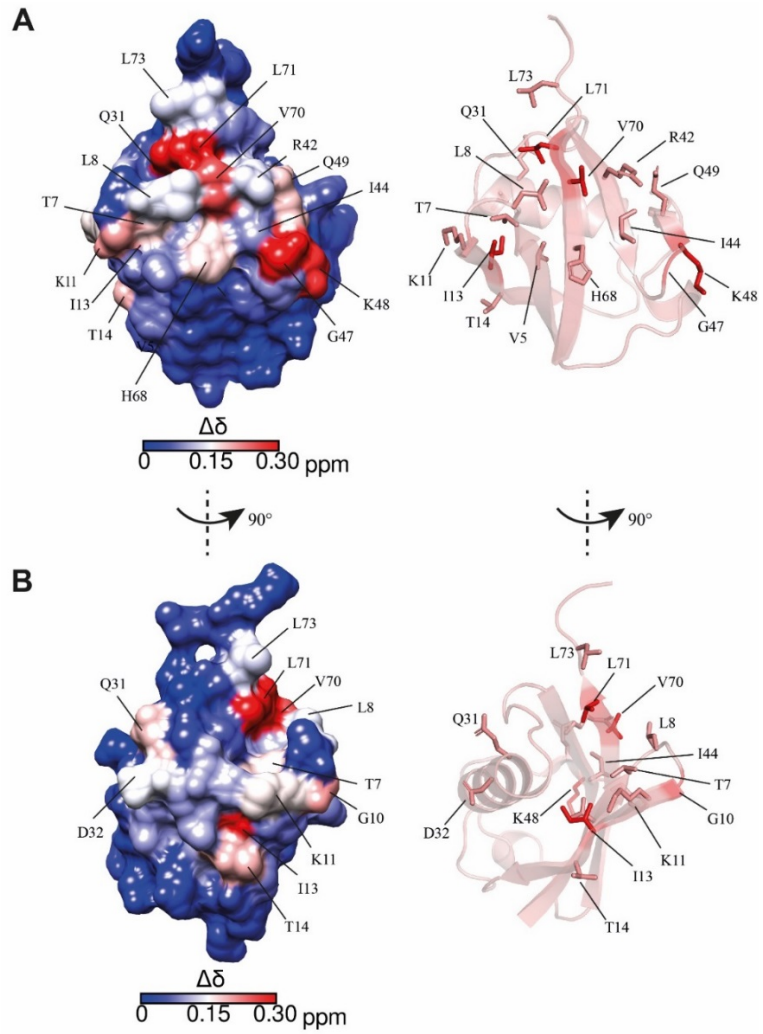
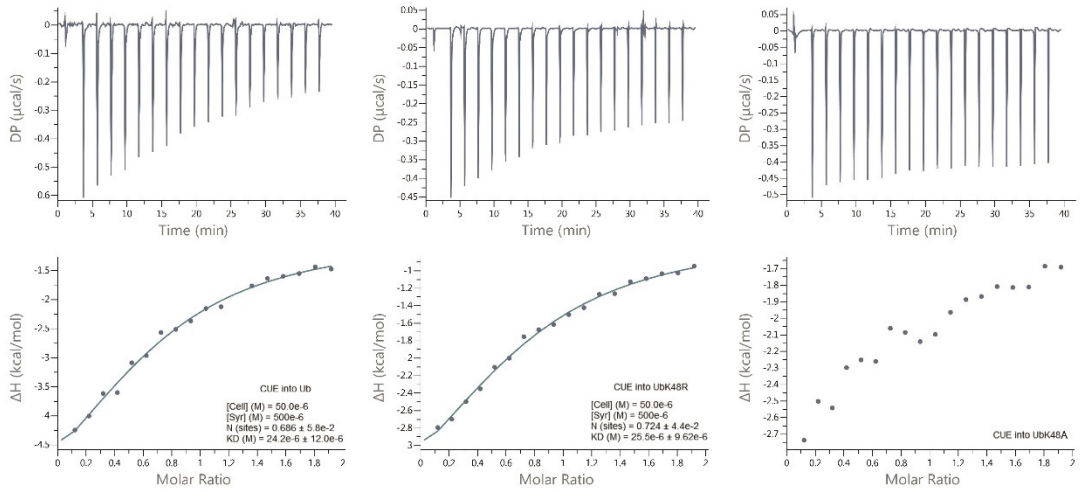


Figure 4.2.40 The perturbed surface of monoUb at maximum saturation (2X) upon binding to the CUE domain. A. The "front" view of the perturbed surface to show the ubiquitin I44 patch. B. The 90° degree view of the perturbed surface. The molar ratio between monoUb and the CUE domain used here is 2:1 (2X). The CSP list at 2X was generated by the subtraction with the chemical shift values from 0X and then mapped to the monoUb surface by CHIMERA. On the right panel of each figure, the key residues with CSP>0.1 are shown as sticks while residues with CSP>0.2 are highlighted in red in the cartoon view by Pymol.

The most perturbed area is the classical I44 hydrophobic patch with some extension including G47 and K48 as shown in Figure 4.2.40A. Despite residue L8 and I44 were only slightly perturbed in the binding event (CSP<0.15), some other residues from this I44 patch including Q49, H68,

V70 and L71 all have higher CSPs upon binding to the CUE domain. Another less perturbed area is next to the I44 patch (Figure 4.2.40B), including residue T7, G10, K11, I13, T14, Q31 and D32.

A



B

Protein pairs	Kd (uM)	N(sites)
CUE (tag-free): monoUb	24.2 ± 12.0	0.69 ± 0.06
CUE (tag-free): K48R	25.5 ± 9.62	0.72 ± 0.04
CUE (tag-free): K48A	undetectable	undetectable

Figure 4.2.41 ITC measurements and the parameter summary of the interaction between the N4BP1 CUE domain and monoUb K48 mutants. A. ITC titration isotherms. B. Parameter table of Kd summarized from A. Data were fitted to a single-site model using the MicroCal PEAQ-ITC Analysis Software following the method described in section 3.2.5.2. Measurement conditions: Syringe: 0.5 mM CUE domain, Cell: 50 µM monoUb or mutants. 18 injections, 2 µL/injection, duration: 2s, spacing time: 120 s, measurement temperature: 25℃.

The importance of ubiquitin K48 for the binding is confirmed by ITC measurements. Ubiquitin mutant K48A completely abolishes the interaction with the CUE domain (Figure 4.2.41). Intriguingly, similar to the WT ubiquitin, the binding affinity between ubiquitin K48R mutant and the CUE domain is unchanged, indicating that the positively charged side chain from K48 plays key roles in the interaction.

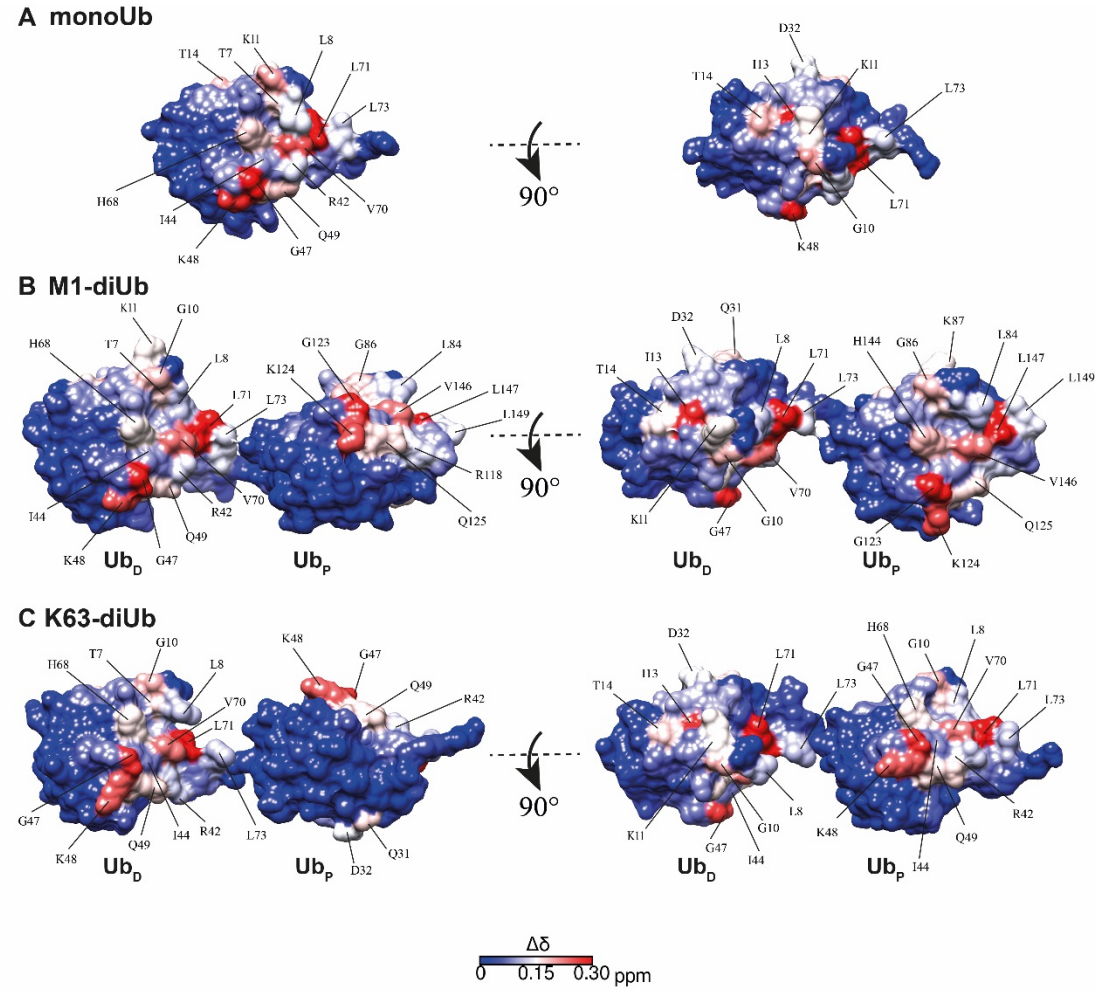


Figure 4.2.42 A comparison of the perturbed surfaces from monoUb (A), M1-diUb (B) and K63-diUb (C) at maximum saturation upon binding to the CUE domain. The molar ratio between monoUb and the CUE domain used here is 2:1 (2X). The molar ratio between diUb (M1- or K63-diUb) and the CUE domain used here is 3:1 (3X). The CSP list for each molecule was mapped to the monoUb surface (PDB: 1UBQ), M1-diUb surface (PDB: 2W9N) and K63-diUb surface (PDB: 3H7P) by CHIMERA respectively. Ub_P: proximal Ub; Ub_D: distal Ub.

The CSPs from M1- and K63-diUb were also mapped to the corresponding surfaces. As seen from Figure 4.2.42, it is clear to find a similar perturbed surface from both ubiquitin moieties. The perturbed surface is also comparable to that of monoUb. Although the two ubiquitin moieties in M1-diUb and K63-diUb have relatively different spatial arrangements, they both adopt an open conformation. This allows the I44 patch from each

ubiquitin moiety well exposed to the solvent. Thus, two molecules of the CUE domain could bind to both ubiquitin moieties from one diUb molecule with an equal chance. This is further supported by the ITC measurements where the stoichiometry between the CUE domain and M1-diUb (or K63-diUb) is 2:1. The affinities of the CUE domain to monoUb or M1- or K63-diUb are also comparable with each other from ITC, supporting that the same binding mode is shared between the three binding pairs.

The above results indicate that the CUE domain does not discriminate among monoUb, M1-diUb and K63-diUb.

In contrast to M1-diUb and K63-diUb, K48-diUb adopts a closed conformation. The crystal structure of K48-diUb reveals that the two I44 patches interact with each other (PDB: 1AAR) (Figure 1.3 in Chapter I), which restricts the I44 patch to interact with binding partners containing UBDs. However, there are several NMR solution structures of K48-diUb in complex with UBDs, which displays a less compacted conformation, allowing the I44 patch more available to the solvent. Here the K48-diUb solution structure taken from the K48-diUb/HHR23A^{UBA} complex (PDB: 1ZO6) was chosen for mapping the CSPs from CUE:K48-diUb to its surface.

As shown in Figure 4.2.43, part of the I44 patch from both proximal and distal ubiquitins was affected upon binding to the CUE domain, but in a less degree of perturbation. The K48 residue in the proximal ubiquitin is occupied to form the K48 linkage, thus the K48 extended surface is blocked. Surprisingly, the K48 extended surface from the distal ubiquitin was also much less involved. This is also supported by the Kd comparison from NMR where the Kd value of K48 in K48-diUb (141.10 μM) is a lot less than the other three ubiquitins (5-8 μM). All the above factors make K48-diUb less accessible to interact with the CUE domain.

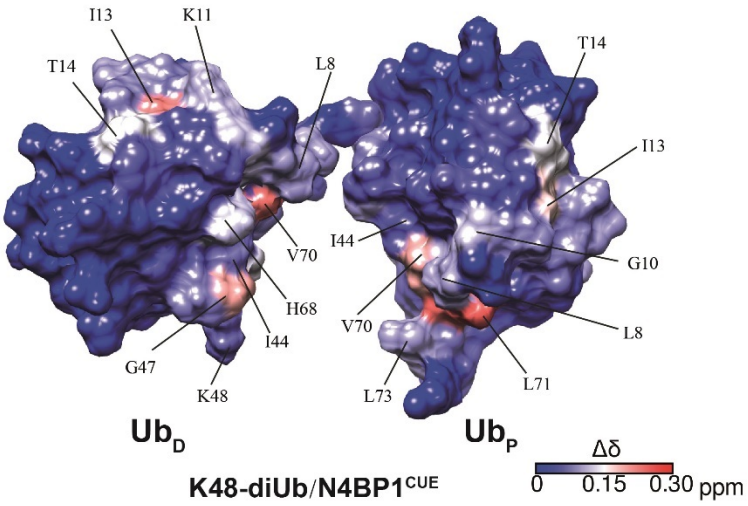


Figure 4.2.43 The perturbed surface of K48-diUb from K48-diUb/N4BP1^{CUE}. The CSPs of K48-diUb upon binding to the N4BP1 CUE domain were mapped to the K48-diUb structure (PDB:1ZO6) by CHIMERA. The molar ratio between K48-diUb and the CUE domain used here is 3:1 (3X). Ub$_P$: proximal Ub; Ub$_D$: distal Ub.

From the NMR titration study in this section, it is clear to conclude that the N4BP1 CUE domain is not in favour of binding to K48-diUb compared to monoUb, M1-diUb or K63-diUb as reflected from the average Kd values summarized in Table 4.2.6 (data taken from Table 4.2.2-4.2.5). Hence, it is further confirmed from the NMR titration studies here that the CUE domain could distinguish K48-diUb from monoUb, M1- and K63-diUb, which is in agreement with the ITC data.

Table 4.2.6 The average dissociation constant (Kd) of ubiquitin upon binding to the CUE domain calculated from NMR chemical shift perturbations

Titration pair	monoUb	M1-diUb	K48-diUb	K63-diUb
CUE	6.44 ± 0.51 μM	14.77 ± 0.27 μM	27.25 ± 0.47 μM	9.56 ± 0.24 μM

Data taken from Table 4.2.2-4.2.5.

The perturbed surface of ubiquitin from each titration pair will be compared with other UBD/ubiquitin complex in the discussion (section 4.2.3.2.3). The CSP data from both ubiquitin and the CUE domain will be used to establish a docking model in the next section, which will help locate the exact binding interface between the two proteins.

4.2.2.4 Molecular docking of the N4BP1 CUE domain to monoUb

Based on the binding studies in section 4.2.2.3, the CUE domain shares the same binding mode to both monoUb and di-ubiquitin chains. To this end, a docking model of the CUE domain in complex with monoUb was generated according to the CSP data collected from NMR titrations (Figure 4.2.44).

In this model, two major features are highlighted: First of all, the FP motif of the CUE domain interacts with residue V70 and L71 from ubiquitin via their hydrophobic side chains. V70 and L71 are part of the ubiquitin I44 hydrophobic patch, although residue I44 from ubiquitin is not involved in the interaction in this model. This is in agreement with the observation from the NMR titrations where I44 was much less perturbed upon binding to the CUE domain.

Another key feature is the polar interaction between residue D893 from the CUE domain and residue K48 from ubiquitin. The negatively charged carboxyl group from the D893 side chain is only 3.8 Å away from the positively charged amino group from the K48 side chain. In section 4.2.2.3.2, the importance of K48 in the interaction has been demonstrated by ITC using the K48A and K48R mutants (Figure 4.2.41). The K48A mutant completely abolishes the interaction between ubiquitin and the CUE domain. On the other hand, the key role of the K48 positively charged side chain is also confirmed by using the K48R mutant. Different from K48A, the K48R mutant retains the same binding affinity as the WT ubiquitin. The side chain of arginine has a similar length to that of lysine. Most important of all, both side chains are positively charged, which very well explains the ability of K48R to interact with the CUE domain.

The N4BP1^{CUE}/monoUb complex will be compared with other CUE/monoUb complexes in the discussion (section 4.2.3.2.3).

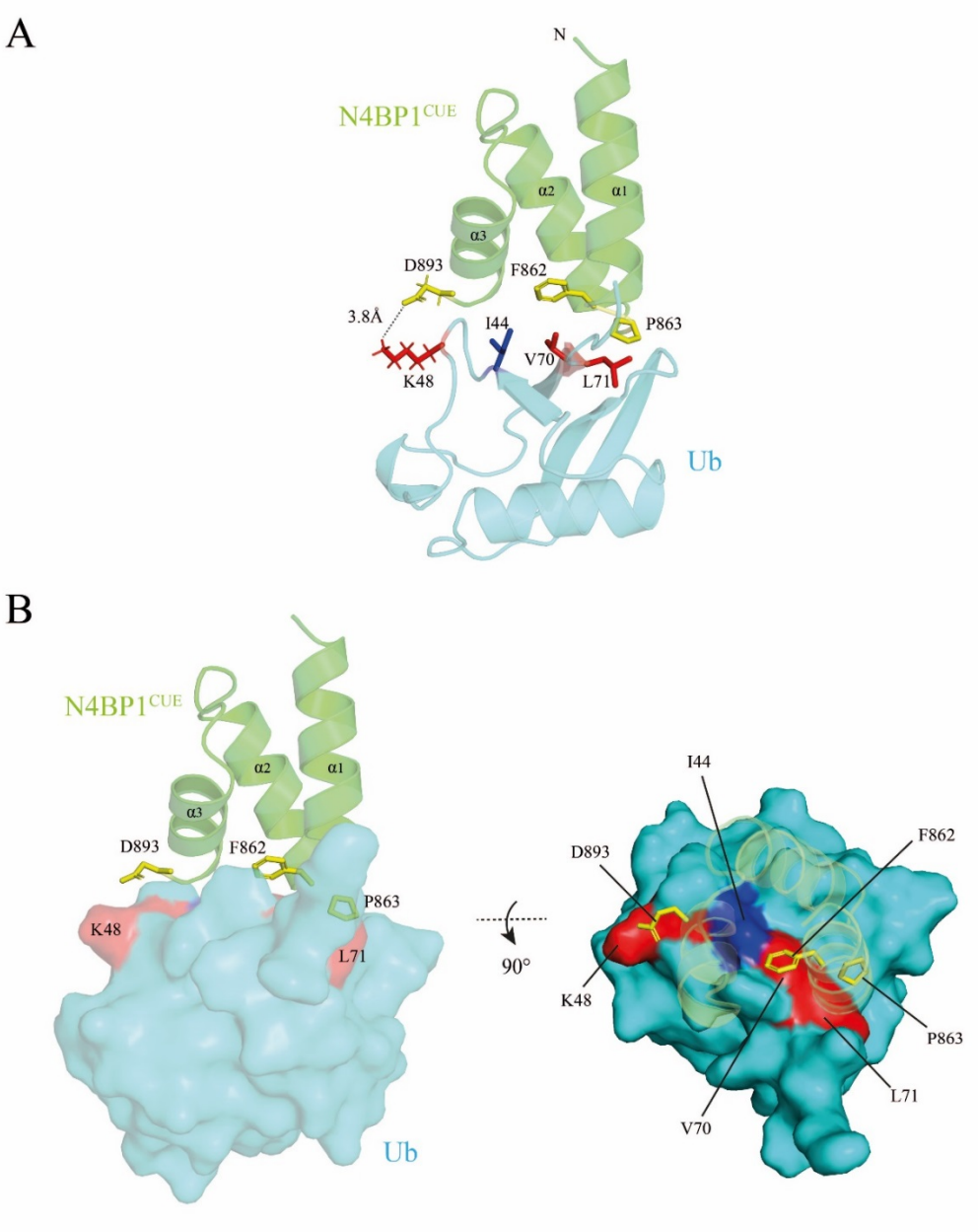


Figure 4.2.44 The docking model of the N4BP1 CUE domain in complex with monoUb. A. Cartoon representation of the model. The hydrogens of residue D893 from CUE and residue K48 from ubiquitin are displayed to help indicate the distance between the two residues. Key residues involved in the interaction are colored in red (for Ub) and yellow (for CUE). B. The ubiquitin is shown in a surface presentation to indicate the binding interface. The N4BP1 CUE domain is coloured in green while the ubiquitin is coloured in cyan. Figure was generated by Pymol.

4.2.2.5 Functional analysis of the N4BP1 RNAse domain

As mentioned in the introduction of this chapter, the pull-down experiments have shown that the N4BP1 RNAse domain contributes to the binding specificity of the CUE domain to M1-diUb and K63-diUb (Figure 4.2.2). This suggests that apart from the RNAse activity, the RNAse domain may have other functions related to ubiquitin binding. This section will be focused on this topic.

4.2.2.5.1 Sequence analysis and structure modelling

In the first instance, a sequence/structural homology searching for the N4BP1 RNAse domain was performed.

Based on the structure prediction of N4BP1 RNAse domain by Phyre2, the top template (c3v33A) is the RNAse domain from the human MCPIP1 (Uniprot: Q5D1E8; reviewed in section 1.4.3). As shown in the sequence alignments, the N4BP1 RNAse domain and the MCPIP1 RNAse domain (both from mouse) are very similar with a 51.5% identity and 72.4% similarity (Figure 4.2.45). This result is also in line with that N4BP1 is evolutionarily close to the MCPIP family from the phylogenetic study (Anantharaman and Aravind, 2006; Figure 1.27).

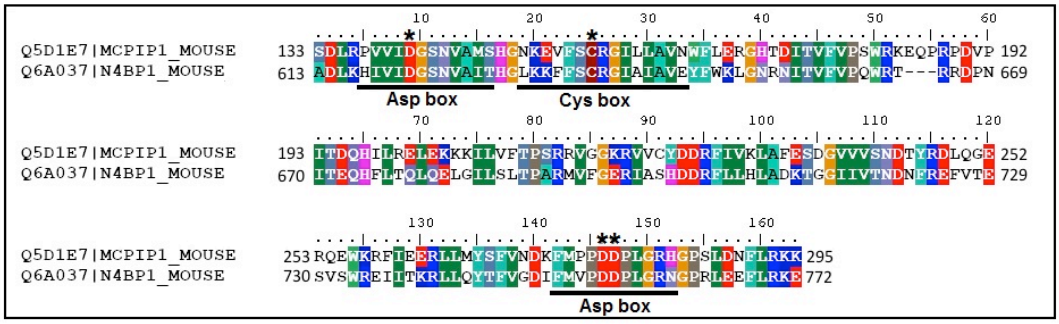


Figure 4.2.45 Sequence alignment of the RNAse domain from mN4BP1 and mMCPIP1. The asterisks (*) indicate the conserved Asp and Cys residues from both domains. The corresponding Asp box and Cys box are indicated. Identical residues are shaded in the same color in pairs while the residues with >80% similarities are also shaded with different color in pairs. Sequence was aligned by EMBOSS Neddle and visualized by Bioedit.

The structure of the RNAse domain from human MCPIP1 has been solved. The negatively-charged pocket which includes the key catalytic residue D141 has been identified to be responsible for the RNAse activity (Figure 4.2.46A) (Xu et al., 2012). D141 is conserved in N4BP1 (The corresponding residue is D623 in human N4BP1 and D621 in mouse N4BP1). In a study using human N4BP1, theD623N mutant lost the RNAse activity (Yamasoba et al., 2019). Here, the *de novo* structure model of mouse N4BP1 RNAse domain (613-774) was predicted by trRosetta. It is quite similar to the MCPIP1 RNAse domain when superimposed together (RMSD=1.748). The negatively-charged pocket is also conserved in both domains (Figure 4.2.46B). It may suggest that both proteins employ the same mechanism to catalyze RNA.

Interestingly, the RNAse domain of MCPIP1 also has the deubiquitinating (DUB) activity towards K48- and K63-linked ubiquitin chains. Residue C157 has been identified as the active Cys site for the DUB activity (Liang et al., 2010). It is no surprise that C157 is also conserved in the N4BP1 RNAse domain (C637 in mouse N4BP1). The DUB signature Asp box and Cys box are also conserved (Figure 4.2.45). This indicates that the N4BP1 RNAse domain might be also involved in ubiquitin-related process.

In the following content, the RNAse construct or the construct containing both the RNAse domain and the CUE domain from N4BP1 will be purified and applied to the biophysical measurements to characterize their binding specificities to monoUb or different diUb chains.

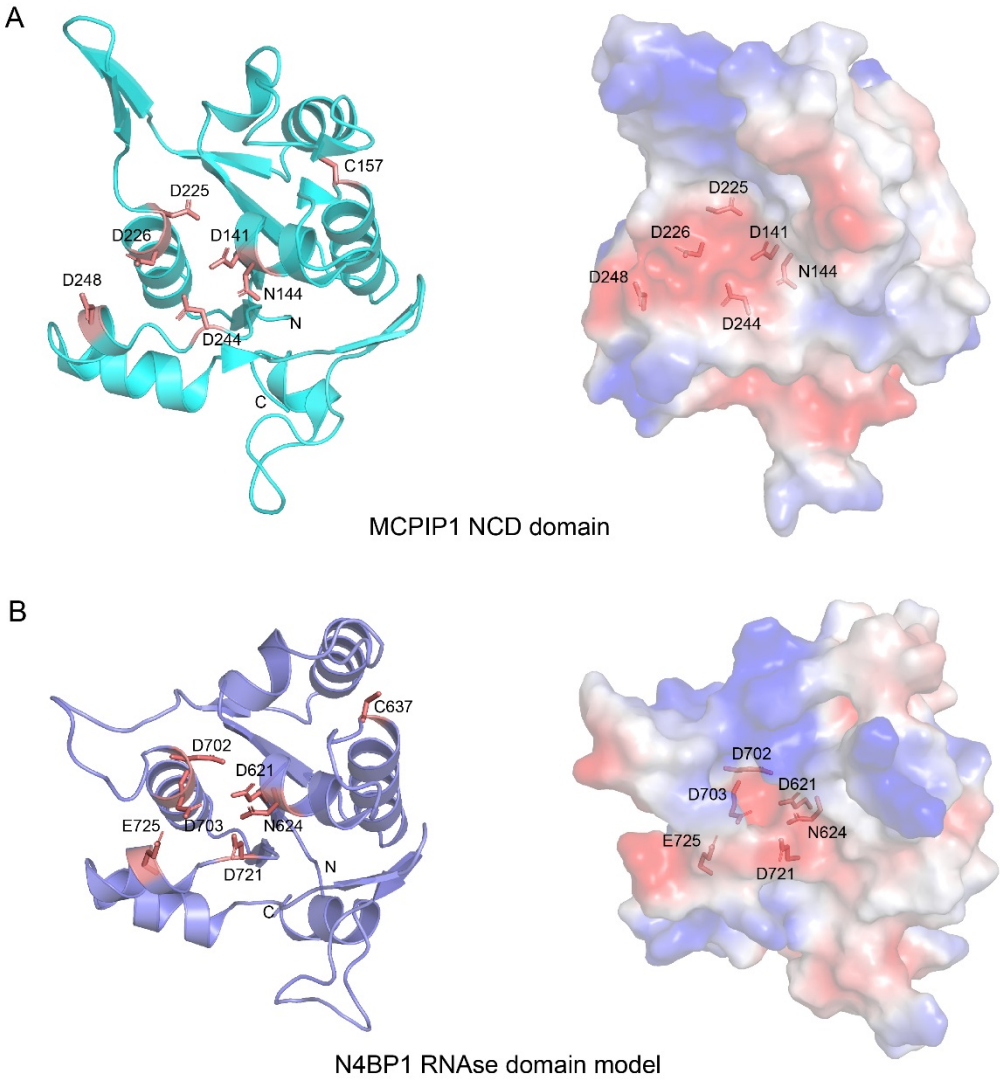


Figure 4.2.46 Structure modelling of the mouse N4BP1 RNAse domain. A. Crystal structure of the human MCPIP1 RNAse domain (PDB: 3V33). B. Structure model of mouse N4BP1 RNAse domain (613-774) by trRosetta. The key residues in the RNAse catalytic center (the negatively-charged pocket) are indicated. The charge distribution on the surface is shown on the right. Positively-charged residues are indicated in blue while negatively-charged residues are indicated in red. Figures were generated by Pymol.

4.2.2.5.2 Protein design and purification

In this section, the short form "RNAse" will be used to represent the RNAse domain (613-774) and "RNAseCUE" to represent the N4BP1 C-terminal construct (613-893) containing both the RNAse domain and the CUE domain. The protein constructs used is summarized in Figure 4.2.47.

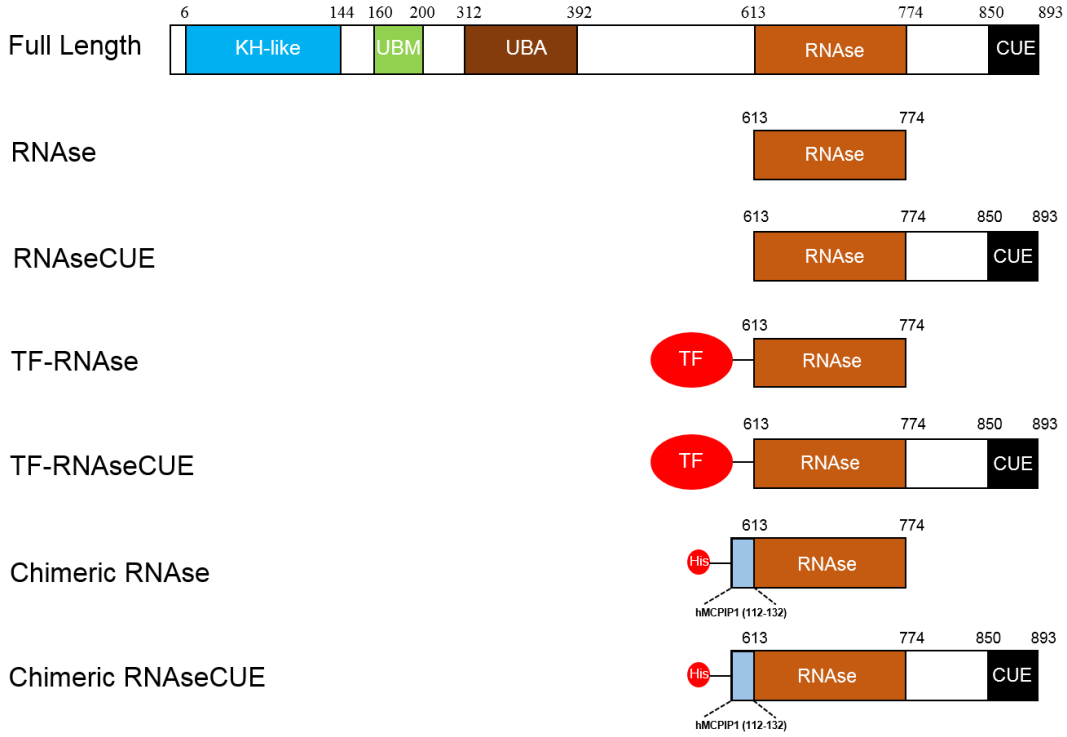


Figure 4.2.47 Domain structures of the RNAse and RNAseCUE constructs. TF: trigger-factor chaperone. His: 6×His tag.

Initial expression tests showed that the expression of RNAse or RNAseCUE was almost all in the inclusion body either using pET47b or pCold-I vector in the *E.coli* bacterial system, indicating the difficulties to make the RNAse domain properly folded and soluble. The pGEX-6P1 vector was also used. Unfortunately, only GST is expressed rather than the fusion protein GST-RNAse or GST-RNAseCUE. The N-terminal boundary of RNAseCUE was also optimized, but again solubility was not improved (data not shown).

To get sufficient soluble protein for biophysical analysis with mono-ubiquitin or di-ubiquitin chains, the below two methods were applied in this study.

1) Construction of the TF-tagged RNAse/RNAseCUE

RNAse and RNAseCUE were cloned into the pCold-TF vector for protein expression. The target protein is fused with a prokaryotic chaperone protein trigger factor (TF, ~50 kDa) in the N-terminus to help increase protein solubility (Figure 4.2.47). The expression of the fusion protein was very strong and a high yield of expressed protein can be obtained with a simple TALON affinity chromatography step (Figure 4.2.48). The tagged protein was further polished by size exclusion chromatography (SEC).

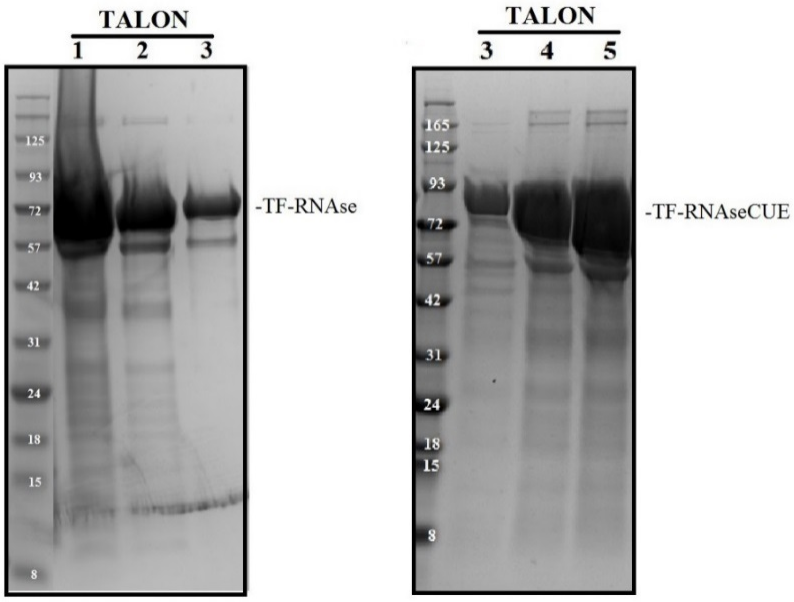


Figure 4.2.48 Purification of TF-tagged RNAse and RNAseCUE by TALON affinity chromatography.

Unfortunately, both RNAse and RNAseCUE precipitated when the trigger factor fusion was cleaved. Therefore, the fusion proteins were applied to the biophysical (ITC/SPR) measurements.

2) Construction of the chimeric RNAse/RNAseCUE

As discussed in 4.2.2.5.1, the N4BP1 RNAse domain shares high homology to that from MCPIP1. Due to the successful purification and crystallization of the human MCPIP1 RNAse domain in prokaryotic *E.coli* expression system (Xu et al., 2012), a sequence optimization of the mouse N4BP1 RNAse domain was performed.

Again, a careful full-length sequence alignment between mouse N4BP1 and human MCPIP1 was conducted, highlighting both similarities and differences (Figure 4.2.49). It has been reported that the human MCPIP1 RNAse domain (112-334) can be crystalized though residues 112-133 are missing in the structural model due to the absence of interpretable electron density (Xu et al., 2012). $hMCPIP1_{112\text{-}132}$ is a short sequence adjacent to the N-terminal of the NCPIP1 RNAse domain which is absent in N4BP1 and might aid solubility for MCPIP1. Therefore, $hMCPIP1_{112\text{-}132}$ was fused to the N-terminus of N4BP1 RNAseCUE (613-893) to create a chimera of the RNAseCUE domain with the aim to increase solubility. The same chimeric protein was also created for the isolated RNAse domain (Figure 4.2.47).

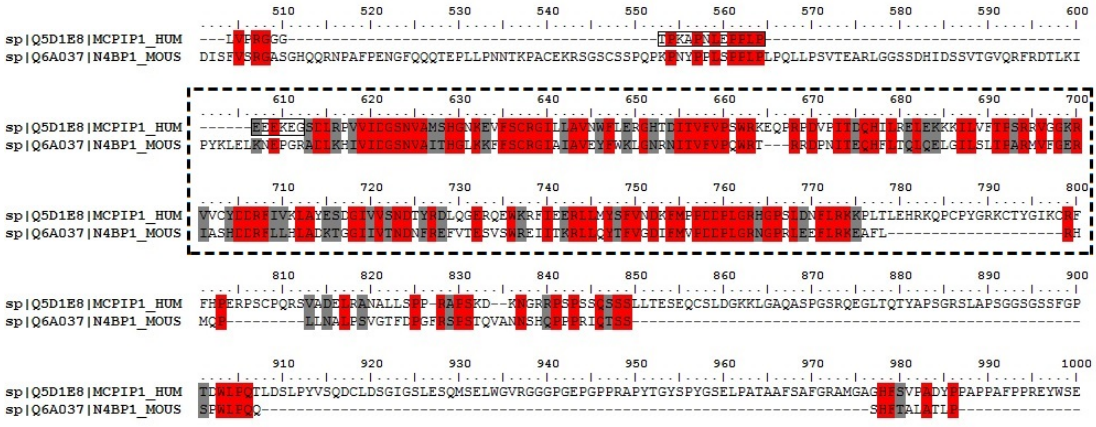


Figure 4.2.49 Full-length sequence alignment between hMCPIP1 and mN4BP1. Note that only part of the alignment is shown here. Dashed box highlights the RNAse domains. Sequence in small black boxes is the adaptor sequence from $hMCPIP1_{112\text{-}132}$. Sequence was aligned by MEGA and visualized by bioedit.

The expression and solubility of both chimeric RNAse and RNAseCUE constructs was strongly improved compared to the corresponding WT construct and can be purified despite large amount of protein are found in inclusion bodies. The purification procedures were optimized into 3 steps. The soluble fraction was first purified by TALON affinity chromatography. Then an anion exchange step using a Mono Q column was performed (Figure 4.2.50). The pure chimeric RNAse or RNAseCUE was eluted at a salt concentration of ~100 mM NaCl. The peak fractions were then applied to analytical size-exclusion chromatography (See details in Figure 4.2.54).

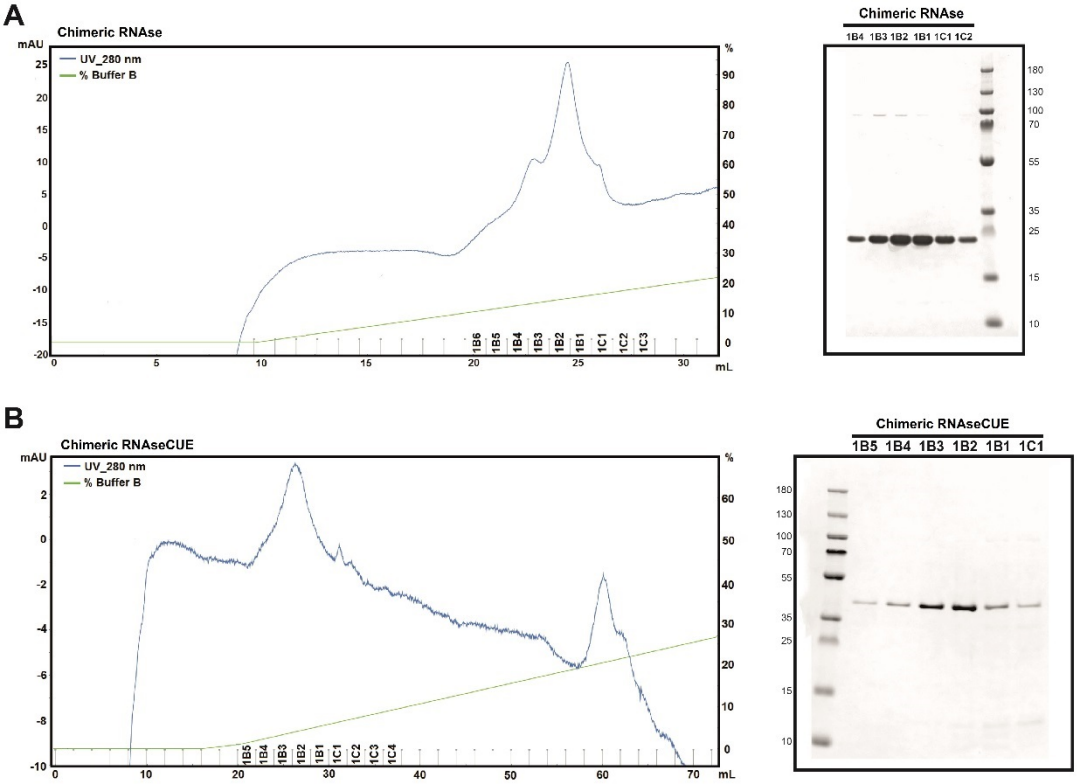


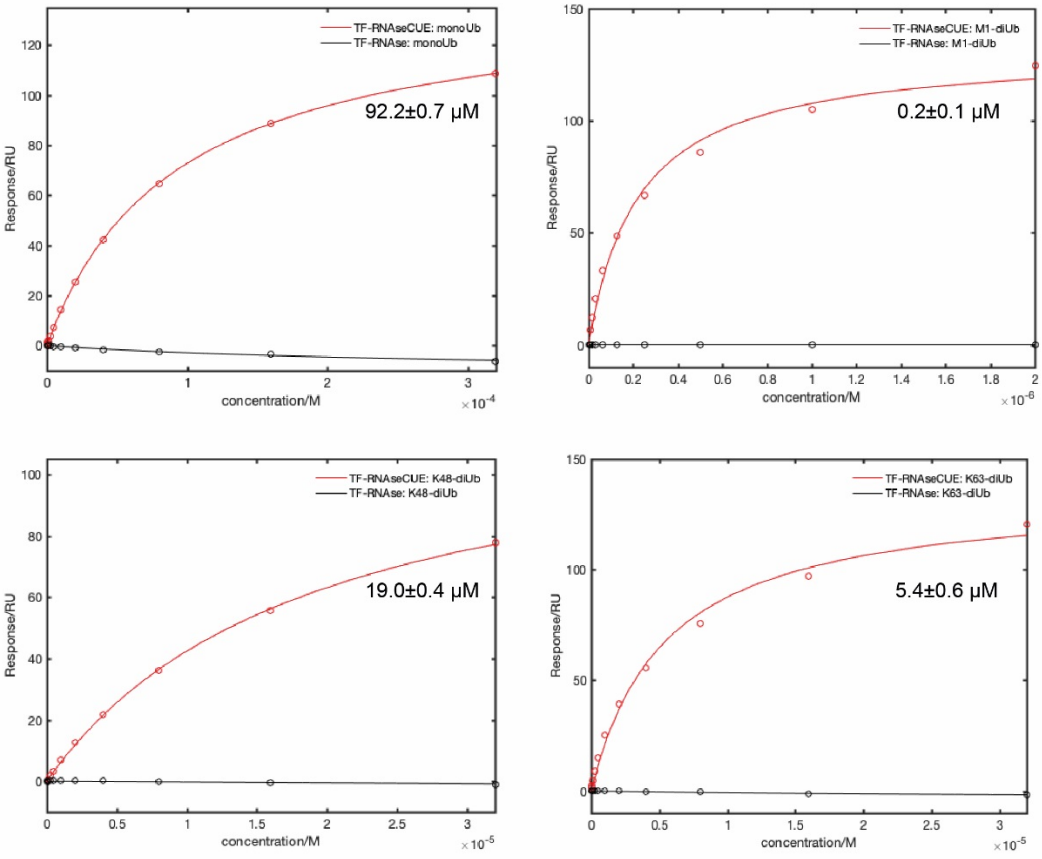
Figure 4.2.50 Purification of chimeric RNAse (A) and RNAseCUE (B) by anion exchange. In each figure, the purification profile is shown on the left while the SDS-PAGE gel of peak fractions is shown on the right. Chimeric RNAse is 23.0 KDa, chimeric RNAseCUE is 36.1 KDa.

4.2.2.5.3 Binding studies

1) SPR measurements reveal the binding preference of RNAseCUE to M1-diUb chains

To characterize the ubiquitin binding properties in a quantitative manner, an affinity screening of RNAse or RNAseCUE was conducted by Surface Plasmon Resonance (SPR). TF-tagged RNAse or RNAseCUE protein was immobilized on a CM5 sensor chip by amine coupling. The analytes (i.e. monoUb or M1/K48/K63-diUb chains in this study) were in the mobile phase and running separately through the chip surface for the interaction.

A



B

Protein	monoUb	M1-diUb	K48-diUb	K63-diUb
TF-RNAse	-	-	-	-
TF-RNAseCUE	92.2±0.7 μM	0.2±0.1 μM	19.0±0.4 μM	5.4±0.6 μM

Figure 4.2.51 Quantitative analysis of TF-tagged RNAse or RNAseCUE interacting with monoUb and M1/K48/K63-diUb by SPR. A. Plots of response at the equilibrium stage from the sensorgrams against the ubiquitin concentration. B. Calculated dissociation constants (K_D) table. Data were fitted by Matlab following the method described in section 3.2.5.3.

The binding constant was calculated using the amplitude of the SPR signal since it was not possible to determine the very fast association and dissociation rates under this condition. The response data points at the equilibrium stage from the sensorgrams (raw data not shown) were plotted against the ubiquitin concentration. The data were fitted to a single-site model to obtain the equilibrium dissociation constant (K_D) as shown in Figure 4.2.51. TF-RNAseCUE displays the highest affinity to M1-diUb with a K_D value of 0.2 μM, then followed by K63-diUb at 5.4 μM. TF-RNAseCUE also binds to K48-diUb, but less effectively (K_D = 19.0 μM) and has the least affinity to monoUb (K_D = 92.2 μM). These results are in line with the observations from the pull-down assays where GST-RNAseCUE prefers binding to M1- and K63-diUb (Figure 4.2.2).

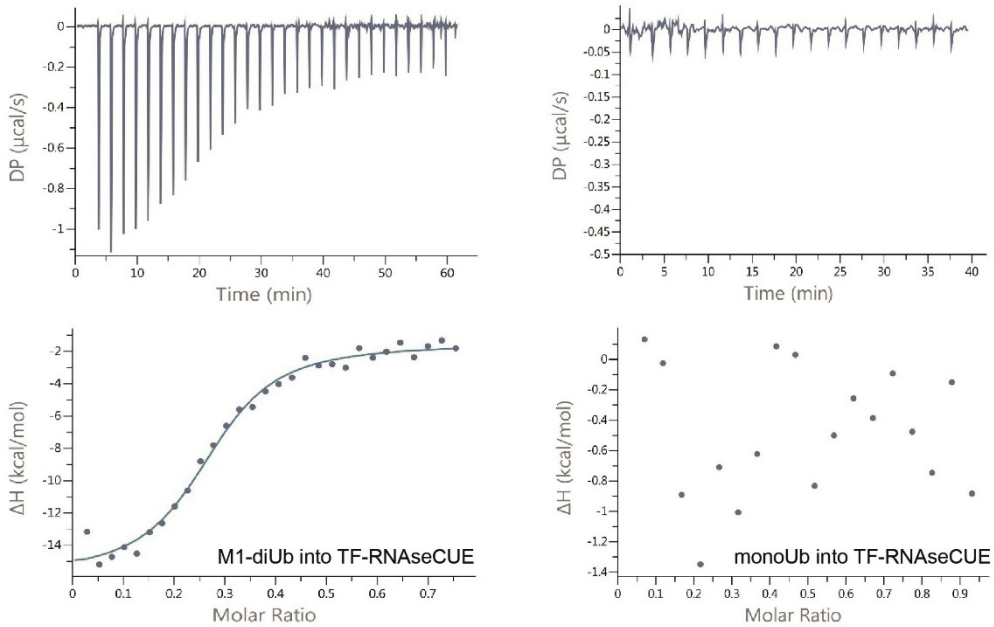
On the contrary, TF-RNAse does not show interactions with any of the four analytes, which indicates the indirect involvement of the RNAse domain in promoting the ubiquitin-binding specificity. To conclude, the C-terminal CUE domain is the sole ubiquitin-binding element in the RNAseCUE construct.

2) ITC measurements confirm the high affinity between RNAseCUE and M1-diUb

To confirm that RNAseCUE binds to M1-diUb with high affinity, ITC measurements were performed. As shown in Figure 4.2.52A, the averaged Kd from three replicates of titrating M1-diUb into TF-RNAseCUE is 2.12 μM. The N value is ~ 0.2, indicating that the ratio between M1-diUb and TF-RNAseCUE is only about 1: 5. If assuming that RNAseCUE still binds to monoUb in the same way as the CUE domain to monoUb at 1:1 ratio, only 20% of the TF-RNAseCUE was involved in the binding event here. This might come from the folding problems from RNAseCUE when the protein is expressed in prokaryotic system as previously mentioned in section 4.2.2.5.2. Nevertheless, the Kd value still indicates a high affinity interaction. As a control, M1-diUb does not interact with TF.

The interaction between TF-RNAseCUE and monoUb was undetectable by ITC (Figure 4.2.52). Based on the Kd value (92.2 μM) measured from SPR, it might be beyond the sensitivity of ITC to detect such weak binding.

A



B

Protein pairs	Kd (μM)	N(sites)
monoUb: TF-RNAseCUE	undetectable	-
M1-diUb: TF-RNAseCUE	2.12±0.40	0.26±0.01

Figure 4.2.52 Quantitative binding measurements of TF-RNAseCUE with M1-diUb and monoUb by ITC. A. The isotherm of titrating 0.5 mM M1-diUb into 0.1 mM TF-RNAseCUE (left), and 0.5 mM monoUb into 50 μM TF-RNAseCUE (right) respectively. B. Parameter table of Kd and stoichiometry summarized from A and B. Data were fitted to a single-site model using the MicroCal PEAQ-ITC Analysis Software following the method described in section 3.2.5.2. Measurement conditions: 18 injections for monoUb into TF-RNAseCUE, 2 μL/injection; 29 injections for M1-diUb into TF-RNAseCUE, 1 μL/injection. Duration: 2s, spacing time: 120 s, measurement temperature: 25℃.

3) ITC measurements confirm that the RNAse domain is not the direct ubiquitin-interacting module

Although the addition of the RNAse domain to the CUE domain significantly increases the specificity to M1-diUb, SPR measurements could not detect any binding between the RNAse domain and ubiquitin (Figure 4.2.51). To further confirm this observation, the ITC measurements were also performed to detect the affinity of TF-tagged RNAse to monoUb or M1-diUb.

As expected, there is no interaction from TF-RNAse with either monoUb or M1-diUb under the sample concentrations as specified in Figure 4.2.53. The intramolecular interaction between the RNAse domain and the CUE domain is also undetectable (Figure 4.2.53C). This further supports that the CUE domain is the sole ubiquitin-interacting module in the C-terminus of N4BP1.

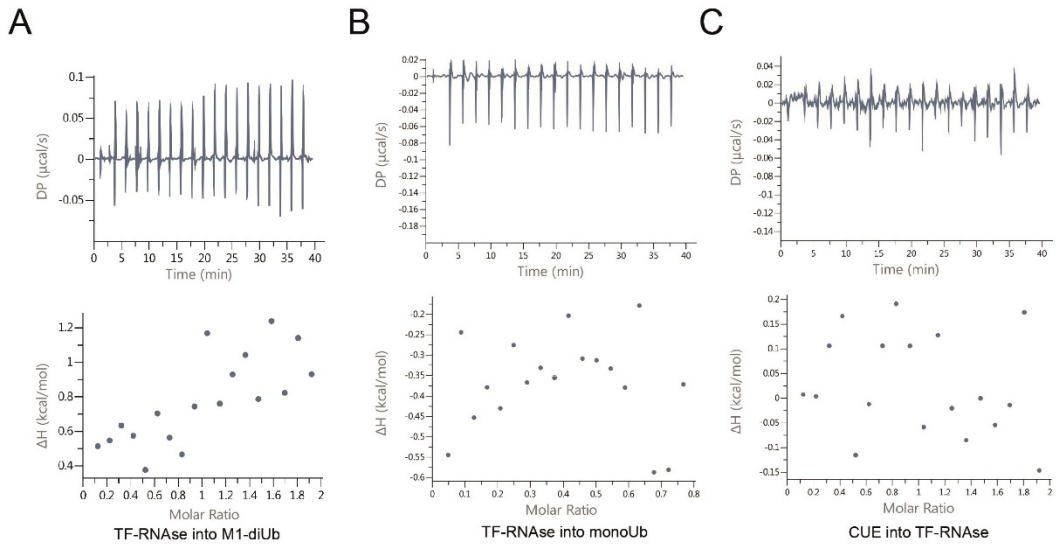


Figure 4.2.53 ITC measurements of TF-RNAse to monoUb, M1-diUb and the CUE domain respectively. A. The isotherm of titrating 0.25 mM TF-RNAse into 25 µM M1-diUb. B. The isotherm of titrating 0.2 mM TF-RNAse into 50 µM monoUb. C. The isotherm of titrating 0.5 mM CUE domain into 50 µM TF-RNAse. Measurement conditions: 18 injections, 2 µL/injection, duration: 2s, spacing time: 120 s, measurement temperature: 25℃.

4.2.2.5.4 Chimeric RNAse and RNAseCUE display oligomeric states

Structural and cell biological studies have revealed that the dimerization of the MCPIP1 RNAse is required to perform its RNAse activity. As introduced in section 1.4.3, the MCPIP1 RNAse domain forms oligomers. The MCPIP1 NTD domain could also interact with the RNAse domain. The interface of NTD-RNAse is partially overlapped with the RNAse dimer interface. Thus, it is proposed that this effect maintains the RNAse domain as dimer instead of oligomers (Yokogawa et al., 2016). Due to the high sequence similarity between the RNAse domains from both MCPIP1 and N4BP1, we hypothesized that the N4BP1 RNAse domain could also form oligomers.

To test this hypothesis, chimeric RNAseCUE was chosen for the analysis. Chimeric RNAseCUE is His-tagged with an additional 20aa-adptor from MCPIP1 attached to the N-terminus, making it a less artificial construct compared to the TF-tagged version. The peak fractions of the corresponding chimeric protein from the anion exchange (as described in Figure 4.2.50) were applied to the analytical size-exclusion chromatography. Several standard proteins with known molecular weight were also analyzed respectively using the same condition as references. As seen from the gel filtration profiles in Figure 4.2.54A, the peak position of both chimeric RNAse (23 KDa) and chimeric RNAseCUE (36.1 KDa) is approximately located in between the Alcohol dehydrogenase peak (150 KDa) and Carbonic Anhydroase peak (29 KDa).

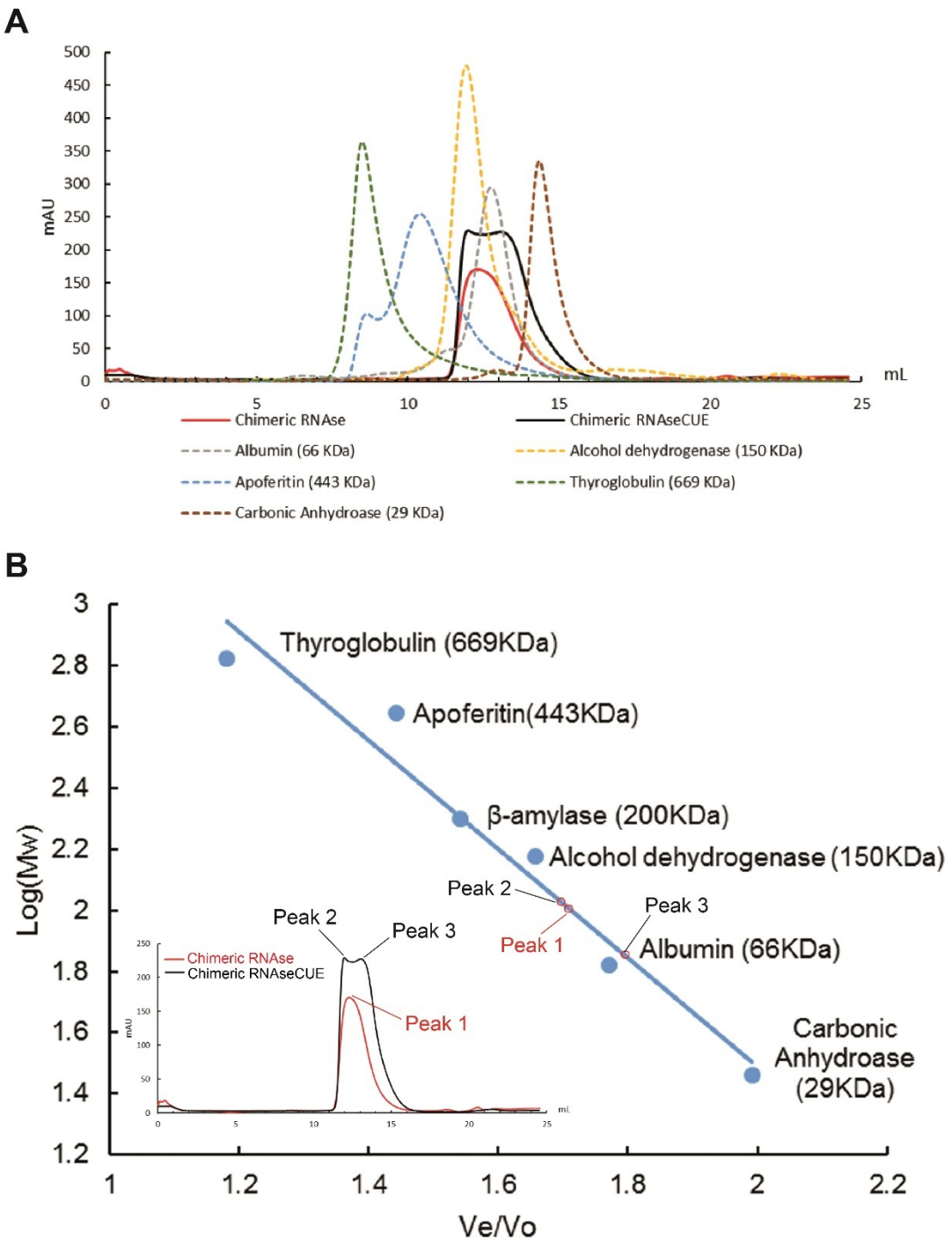


Figure 4.2.54 Analytical size-exclusion chromatography reveals oligomeric states of chimeric RNAse and chimeric RNAseCUE. A. The analytical size-exclusion profiles of the chimeric proteins as well as the standard proteins. The Mw of the standard protein is indicated. The method followed what's described in section 3.2.5.1. B. The calibration curve of the gel filtration column. The column used here is Superose™ 12 10/300L (GE Healthcare). Ve is the elution volume of the corresponding protein peak. Vo is

the void volume of the column (7.2 mL). The curve was linearly fitted to the equation y=-1.7788x+5.0463 where x is the Ve/Vo ratio and y is the Log10 value of the protein molecular weight. The gel filtration profiles of the two chimeric proteins are shown in the left corner. The Ve/Vo of each peak was applied to the fitted equation correspondingly to estimate the molecular weight as indicated by the red circles in the calibration curve.

The molecular weight of the two chimeric proteins was calculated according to the calibration curve derived from the gel filtration profiles of the standard proteins (Figure 4.2.54B). The elution volume/void volume ratio (Ve/Vo) of the 3 peaks of interests (i.e. Peak 1 from chimeric RNAse; Peak 2 and 3 from chimeric RNAseCUE) was calculated respectively from the corresponding gel filtration profile. Then the molecular weight of the protein from each peak was estimated by applying its Ve/Vo ratio to the calibration curve (Figure 4.2.54B). The estimated molecular weight is summarized in Table 4.2.7.

From the results, the single peak of the chimeric RNAse corresponds to a molecular weight of a tetramer. There are two peaks close to each other in the chimeric RNAseCUE profile where Peak 2 corresponds to a tetramer and Peak 3 indicates a dimeric state. This shows that chimeric RNAseCUE exists in an equilibrium of the dimeric and tetrameric species, which cannot be separated by gel filtration.

Table 4.2.7 The molecular weight estimation of the protein peaks from the analytical SEC profiles of chimeric RNAse and chimeric RNAseCUE

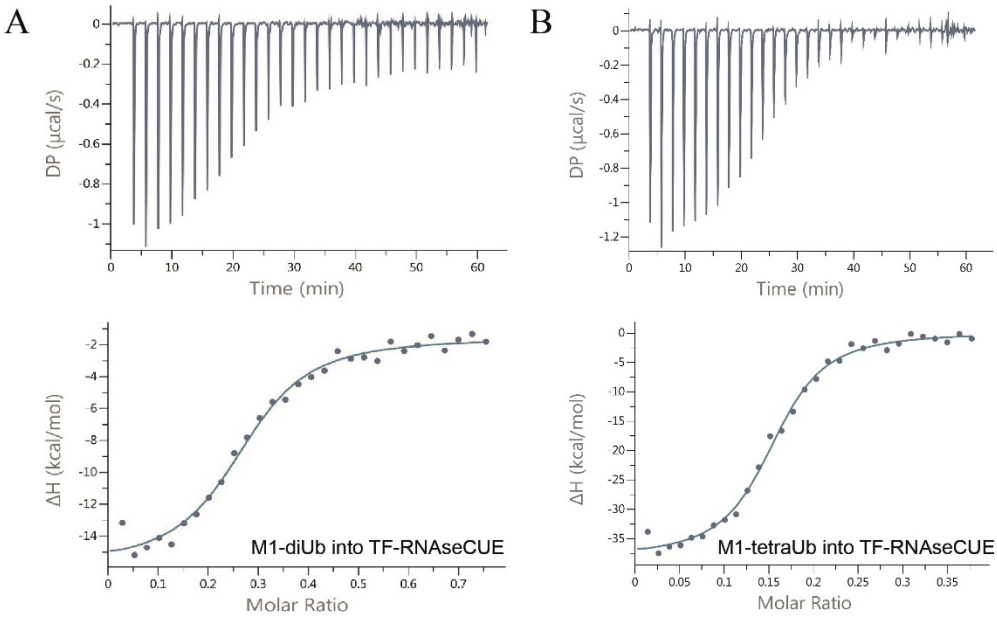
Protein	Monomer Mw.	Peak	Calculated Mw.	Oligomeric states
Chimeric RNAse	23 KDa	Peak 1	102.91 KDa	tetramer
Chimeric RNAseCUE	36.1 KDa	Peak 2	121.37 KDa	tetramer
		Peak 3	65.66 KDa	dimer

In summary, the gel filtration profiles of both chimeric RNAse and chimeric RNAseCUE indicate the oligomeric state of the N4BP1 RNAse domain present in the solution.

Based on the above experimental evidence, a hypothesis was proposed: the oligomerization of the RNAse domain brings the C-terminal CUE domains together in close proximity, which increases the chance for linear ubiquitin chains to recognize more CUE molecules at a time. The high affinity of the interaction between the RNAseCUE construct and the linear ubiquitin chains is achieved not only by the direct binding of the CUE domain but also from the appropriate structural arrangements of the RNAse domains.

Here, to verify this hypothesis, the interaction of TF-RNAseCUE with M1-diUb or M1-tetraUb was compared by ITC measurements (Figure 4.2.55). M1-tetraUb was titrated into TF-RNAseCUE, resulting in a Kd value of 0.59 μM, which is about 4 times tighter than the affinity between TF-RNAseCUE and M1-diUb (2.12 μM) from the study in section 4.2.2.5.3. The increased affinity further supports that the oligomeric RNAse domain contributes to the increased local concentration of the N4BP1 C-terminus, which makes the whole RNAseCUE construct prefer to interact with longer linear ubiquitin chains.

A structural model of dimer-induced specificity of N4BP1 for linear ubiquitin chains will be proposed and discussed in section 4.2.3.3.



C

Protein pairs	Kd (µM)	N(sites)
M1-diUb: TF-RNAseCUE	2.12±0.40	0.26±0.01
M1-tetraUb: TF-RNAseCUE	0.59±0.09	0.15±0.002

Figure 4.2.55 Quantitative binding measurements of TF-RNAseCUE to M1-diUb and M1-tetraUb by ITC. A. The isotherm of titrating 0.5 mM M1-diUb into 0.1 mM TF-RNAseCUE. The data was taken from Figure 4.2.51. B. The isotherm of titrating 0.25 mM M1-tetraUb into 0.1 mM TF-RNAseCUE. C. Parameter table of Kd and stoichiometry summarized from A and B. Data were fitted to a single-site model using the MicroCal PEAQ-ITC Analysis Software following the method described in section 3.2.5.2. Measurement conditions: 29 injections, 1 µL/injection, duration: 2s, spacing time: 120 s, measurement temperature: 25℃.

4.2.3 Discussion

4.2.3.1 Structure comparison of the CUE domain with other UBDs

In section 4.2.2.2, the structure model of the N4BP1 CUE domain was generated based on the backbone and sidechain assignments by solution NMR. The N4BP1 CUE domain adopts a 3-helix bundle structure which is similar to the canonical CUE family. The structure is also in agreement with the *de novo* structure prediction by trRosetta. The CS-ROSETTA model and the trRosetta model can be very well aligned (RMSD=0.895) (Figure 4.2.56). Thus, the CS-ROSETTA model allows the further characterization of the ubiquitin-binding properties of the N4BP1 CUE domain.

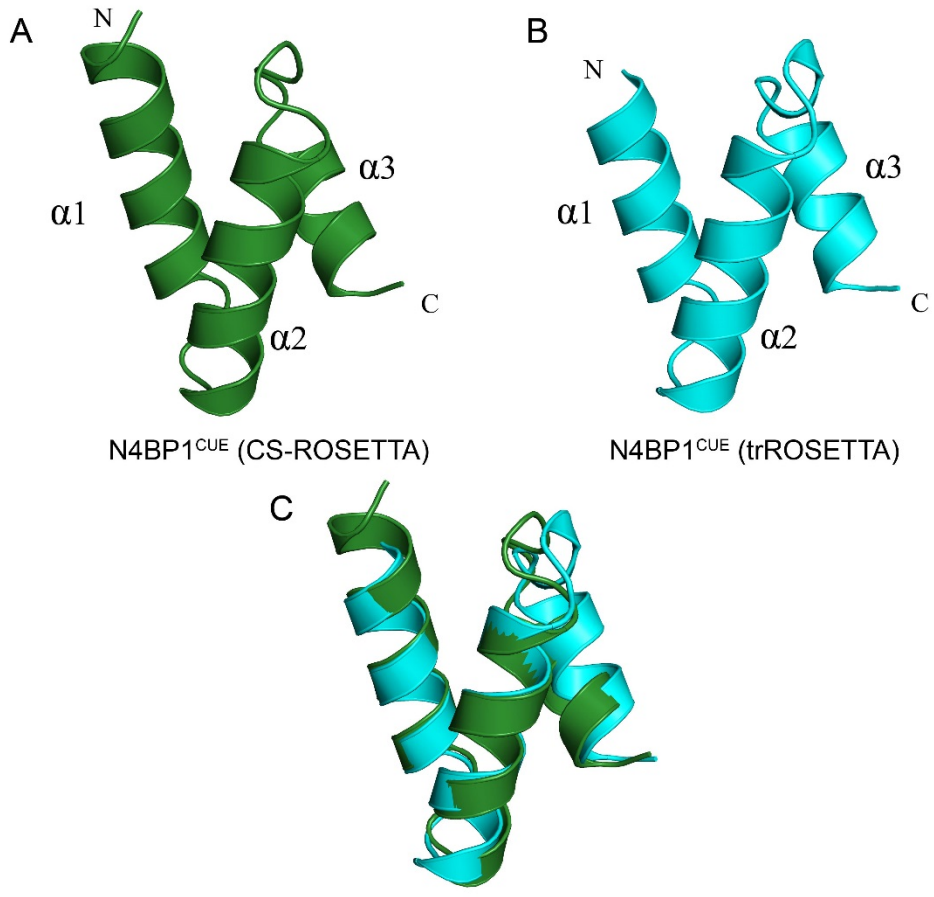


Figure 4.2.56 Structure superimposition of the CS-ROSETTA model and trRosetta model. A. Structure model of CUE by CS-ROSETTA. B. Structure model by trRosetta. C. Structure superimposition of the two models. Figures were generated by Pymol.

A protein structure comparison via the DALI server was performed using this CS-ROSETTA model as an entry. Among the several top results, the CTD domain of MCPIP1 was found (PDB: 2N5L). Most of the other top structures only share similar 3-helix fold with the N4BP1 CUE domain, but lack the conserved FP motif or have no related ubiquitin-binding properties. Thus, they will not be discussed here.

The structure of N4BP1 CUE domain and MCPIP1 CTD domain can be aligned as shown in Figure 4.2.57 (RMSD=3.308). The main difference between the two domain is that the helix α3 from $MCPIP1^{CTD}$ is longer than $N4BP1^{CUE}$. Nevertheless, the FP motifs from both domains are conserved and located in the loop between helix α1 and α2. Although the structure of $MCPIP1^{CTD}$ is available, the classification of this domain is yet unclear. The function of $MCPIP1^{CTD}$ also remains further investigation. However, considering that the full length MCPIP1 has three domains that are similar to N4BP1 (See the discussion in section 4.2.3.4; see Appendix III for the full sequence alignments), the reported function of MCPIP1 in immune responses can be helpful for the functional investigation of N4BP1 in the future study.

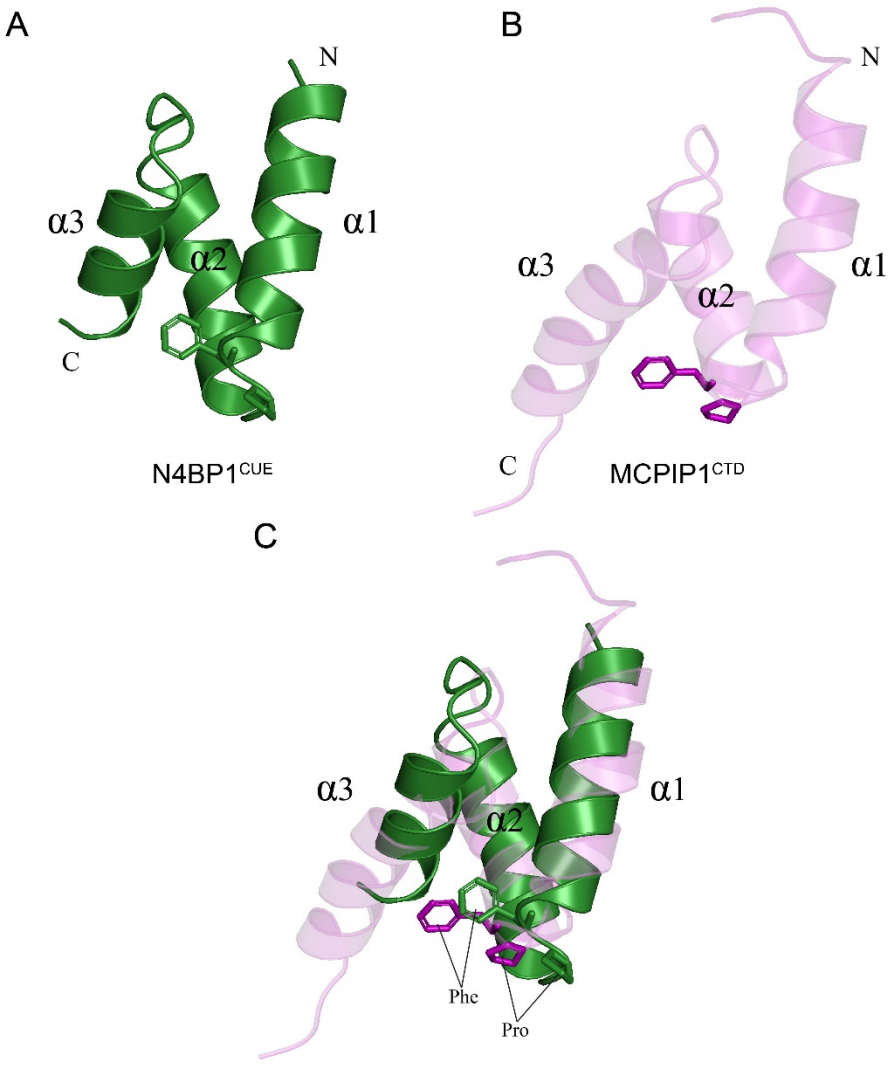


Figure 4.2.57 Structure superimposition of N4BP1^{CUE} and MCPIP1^{CTD}. A. Structure model of N4BP1^{CUE}. B. Solution NMR structure of MCPIP1^{CTD} (PDB: 2N5L). C. Structure superimposition of the two structures. Figures were generated by Pymol.

Apart from the CTD domain from MCPIP1, several other UBDs can be also found, including the CUE domain from gp78 (also called AMFR or RNF45) (PDB: 2EJS) and CUE2 (PDB: 1OTR), the UBA domain from p47 (PDB: 1V92) etc (Figure 4.2.58). Interestingly, although sharing high sequence similarity with the N4BP1 CUE domain, the CUBAN domain from KHNYN (PDB: 2N5M) cannot be found via DALI search. A careful structural comparison between the N4BP1 CUE domain and these UBDs will be carefully discussed below respectively.

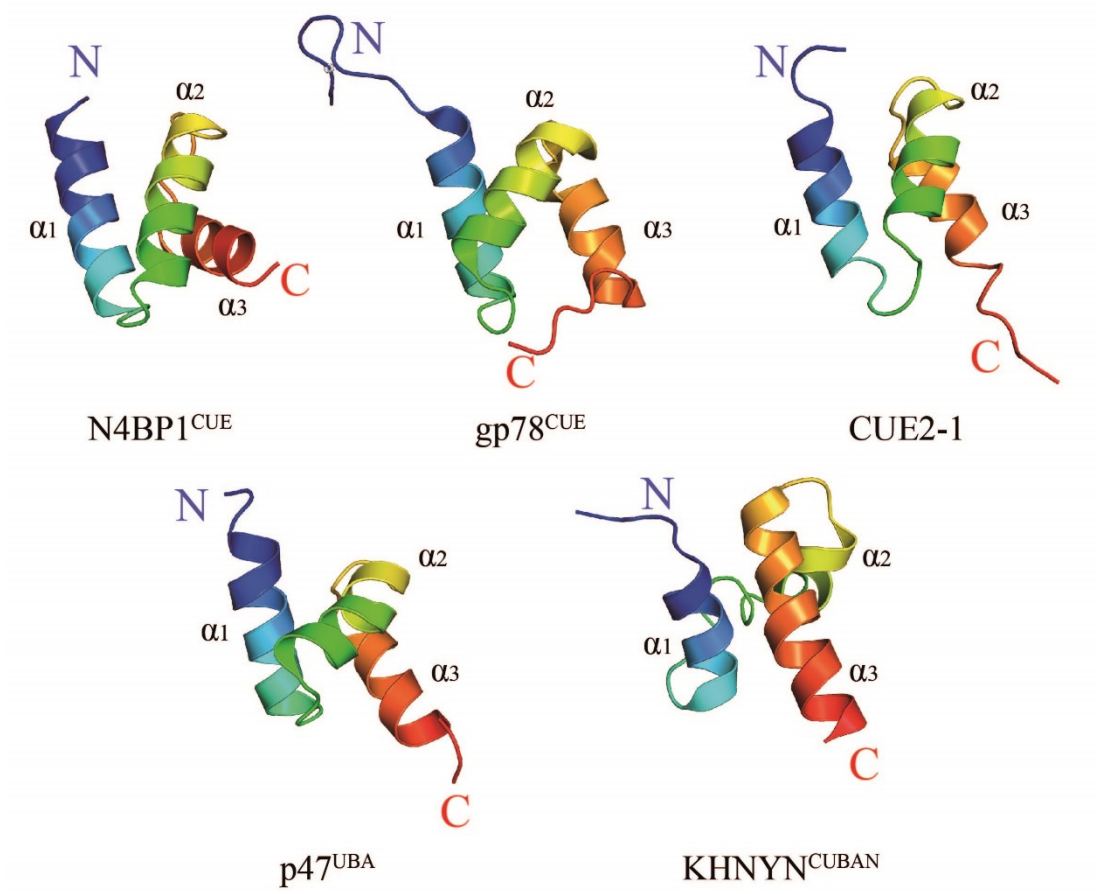


Figure 4.2.58 Structures of the N4BP1 CUE domain and representative CUE domain, UBA domain and CUBAN domain. All the α1 helices are colored from dark blue to light blue. All the α2 helices are colored from green to yellow. All the α3 helices are colored from orange to red. Figures were generated by Pymol using the structure model of N4BP1^{CUE} (from this thesis), gp78^{CUE} (PDB: 2EJS), CUE2-1 (PDB: 1OTR), p47^{UBA} (PDB: 1V92), KHNYN^{CUBAN} (PDB: 2N5M).

Due to poor sequence similarity, the N4BP1 CUE domain could not be aligned with either the gp78^{CUE} or CUE2-1 domain by a simple Pymol sequence alignment. Then a pairwise alignment was performed manually as shown in Figure 4.2.59. The alignment of the first helix (α1) from both pairs allows matching the three helical bundles in good agreement. According to the sequence alignments from Figure 4.2.3, the FP motifs are conserved. The

structural comparison here further confirmed the conserved position of the FP motifs among the three domains.

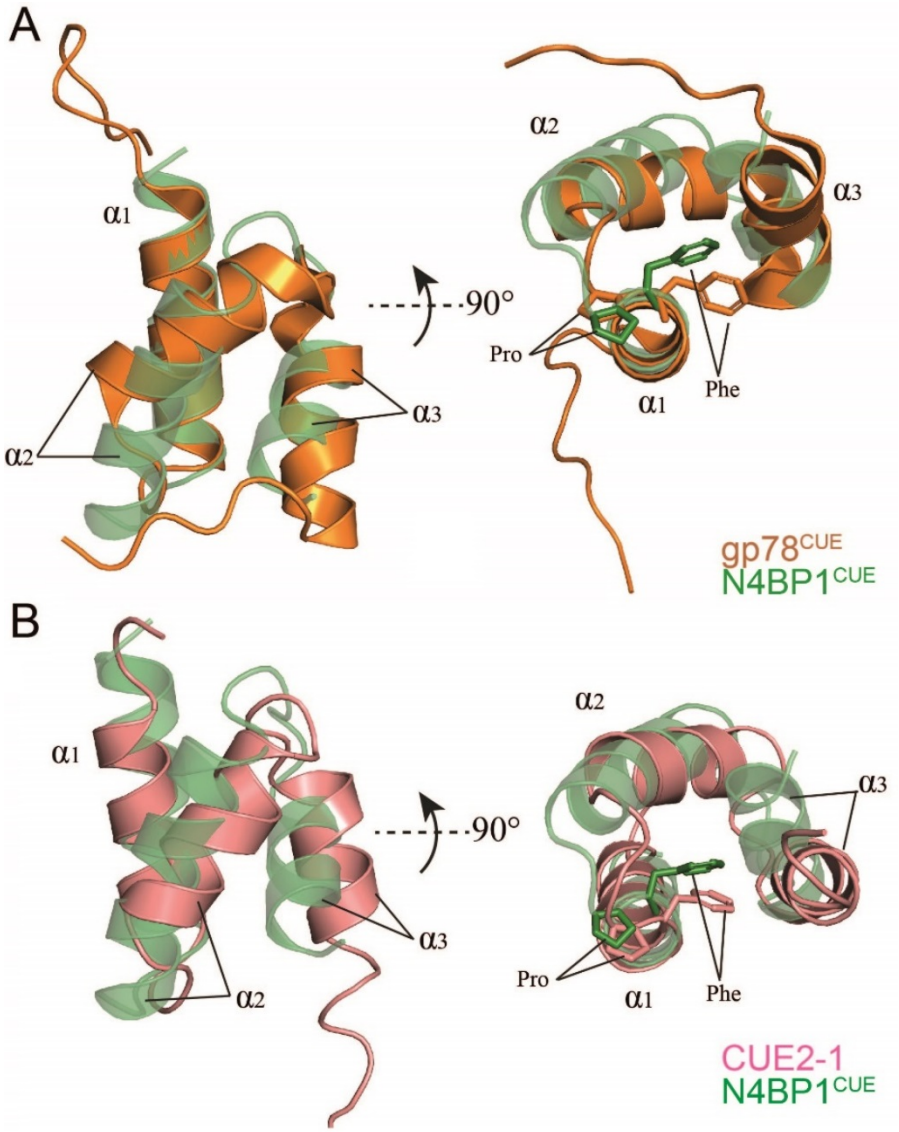


Figure 4.2.59 The structure comparison of the N4BP1 CUE domain with gp78^{CUE} (A) and the CUE2-1 domain (B). The first helix (α1) from both structure pairs are manually aligned together. FP motifs are highlighted in both structures. Figures were generated by Pymol using the structure model of N4BP1^{CUE} (from this thesis), gp78^{CUE} (PDB: 2EJS) and CUE2-1 (PDB: 1OTR).

For the comparison between N4BP1^{CUE} and p47^{UBA}, surprisingly, helix α3 from p47^{UBA} is aligned with helix α1 from N4BP1^{CUE} by Pymol according to sequence similarity (Figure 4.2.60A). Manual superimposing the first helix

(α1) from both domains results in a complete misalignment of the second helix (α2) as shown in Figure 4.2.60B. The comparison highlights the difference in both sequence and helix fold between the two domains.

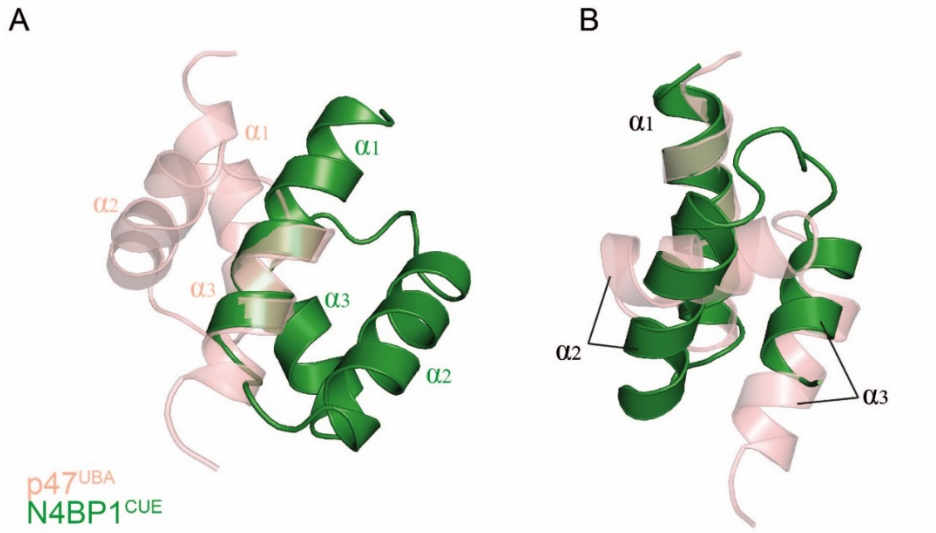


Figure 4.2.60 The structure comparison of the N4BP1 CUE domain with the p47 UBA domain. A. Structure superimposition by sequence alignment by Pymol. B. Structure superimposition by manually aligning α1 from both domains together. Figures were generated by Pymol using the structure model of $N4BP1^{CUE}$ (from this thesis) and $p47^{UBA}$ (PDB: 1V92).

As mentioned previously in this chapter, the top template for the sequence-based structural prediction of the N4BP1 CUE domain by Phyre2 is the KHNYN CUBAN domain (PDB: 2N5M). A simple sequence alignment indicates that they share ~ 33% identity and ~ 46% similarity (Figure 4.2.61A). However, as seen from Figure 4.2.58, it is obvious that the structures differ substantially. The relative locations of helix α2 and α3 from both proteins display a different spatial arrangement. A superimposition of the two domains demonstrates that the structures have separate folds which cannot be aligned (Figure 4.2.61B). This observation is in line with the different binding preferences of the two domains to NEDD8 and ubiquitin.

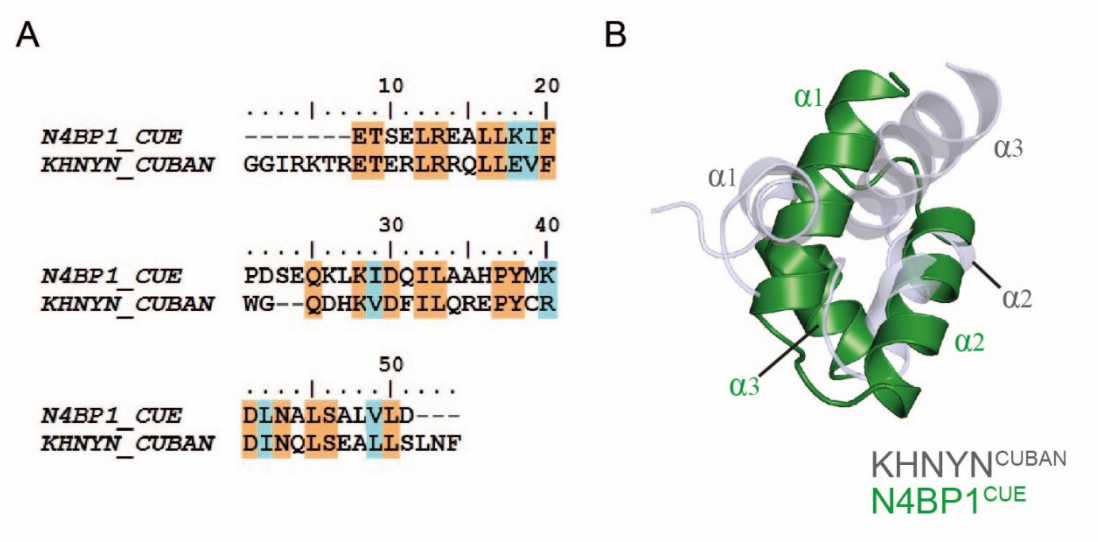


Figure 4.2.61 Sequence and structure comparisons of N4BP1^{CUE} with KHNYN^{CUBAN}. A. Sequence alignments between the two domains. Identical residues are shaded in orange. The residues with >80% similarities are shaded in cyan. Sequence alignments were done by EMBOSS Needle and visualized by bioedit. B. The superimposition of the two domains. Figures were generated by Pymol using the structure model of N4BP1^{CUE} (from this thesis) and KHNYN^{CUBAN} (PDB: 2N5M).

From the above structural comparisons, despite the high sequence similarity between the N4BP1 CUE domain and the KHNYN CUBAN domain, the N4BP1 CUE domain is structurally conserved with the members from the canonical CUE family. It is particularly highlighted from the conservation of the FP motif. Hence, this domain is grouped into the CUE domain family.

At the middle stage of this project, the research group who solved the KHNYN CUBAN domain also published the human N4BP1 CUE structure by solution NMR and named it CoCUN (Nepravishta et al., 2019). However, despite the 3-helix arrangement, Nepravishta and colleagues claimed that CoCUN has a large part of random coils and shares similar protein fold with KHNYN^{CUBAN}, which is not in agreement with the finding in this thesis. By carefully looking at the NMR procedure, we found that all the assignments of the carbon atoms (from both backbone and sidechains) from CoCUN were

absent in the chemical shift list. Thus, it is questionable that the CoCUN structure provided by this group is accurate. Since the CoCUN structure is not deposited in the PDB database, it is difficult to compare the results of this work with their structure in more detail. The CoCUN/ubiquitin complex structure was also present by that group. It will be discussed in section 4.2.4.2.3.

4.2.3.2 The ubiquitin-binding specificity of the N4BP1 CUE domain

In addition to the structural study of the N4BP1 CUE domain, investigating the interaction between the N4BP1 CUE domain and ubiquitin is another key topic. Thus, the binding studies were performed by multiple biophysical approaches including SPR, ITC and NMR titrations in section 4.2.2.1 and 4.2.2.3.

Quantitative analysis by SPR shows that the GST-tagged CUE domain has high affinity to M1-diUb compared to other types of ubiquitin. However, it was later demonstrated that the dimeric GST has an artificial effect on the affinity. Thus, the ITC technique was chosen to be an ideal approach which allows using the tag-free CUE domain for the characterization. The ITC results demonstrate that the N4BP1 CUE domain has lower binding affinity to K48-diUb compared to monoUb, M1-diUb and K63-diUb. The affinity of CUE to the latter three ubiquitins is comparable to each other.

To further characterize the ubiquitin-binding mode, NMR titrations were performed to locate the perturbed surfaces from both the CUE domain and ubiquitin, which allows the below analysis for surface comparison.

4.2.3.2.1 Perturbed surface comparison – from the CUE domain side

As demonstrated in section 4.2.2.3.1, by mapping the CSPs from NMR titrations to the CUE surface, the main perturbed area is within the groove between helix α1 and α3. The key residues from this perturbed area was compared with those involved in ubiquitin binding from the well-characterized CUE2-1 domain (PDB: 1OTR) and gp78 CUE domain (PDB: 2LVO).

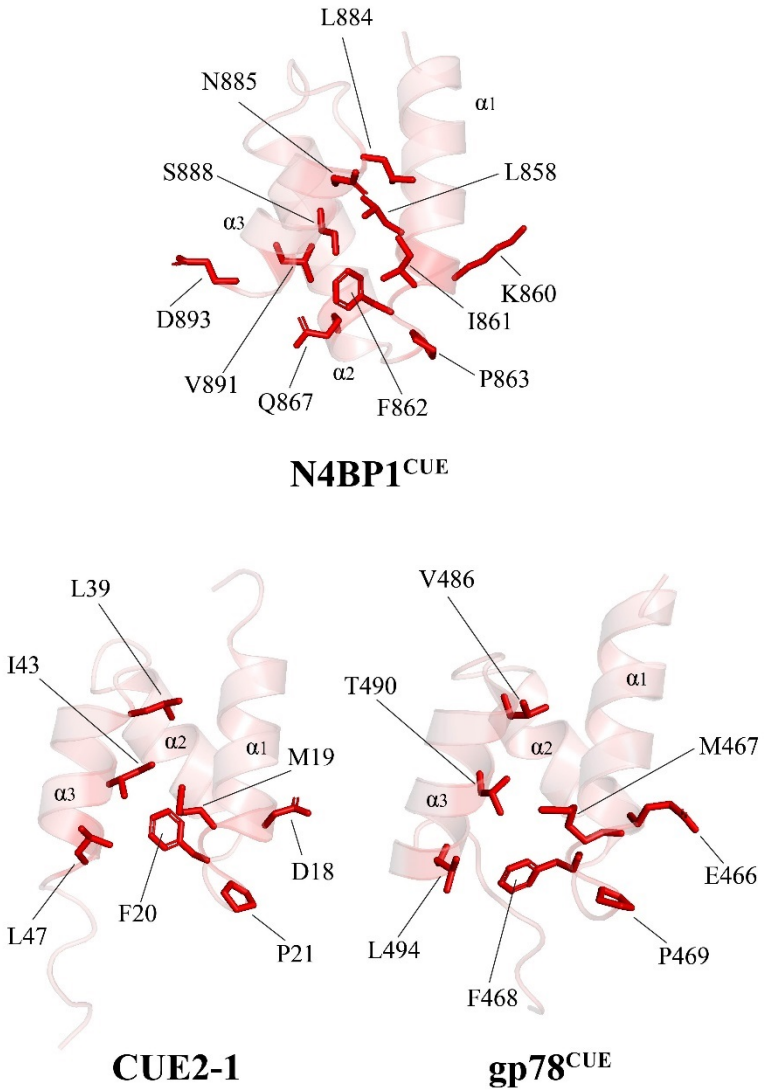


Figure 4.2.62 The comparison between the key perturbed residues of N4BP1 CUE domain, CUE2-1 domain and gp78 CUE domain. Key residues with strong CSPs (>0.15) from N4BP1 CUE domain and residues involved in ubiquitin-binding in CUE2-1 domain and gp78 CUE domain are highlight as red sticks. The CUE2-1 structure is generated by Pymol using PDB: 1OTR. The gp78 CUE domain structure is generated by Pymol using PDB: 2LVO.

From this perspective, it appears that all three CUE domains share a similar binding surface for the interaction (Figure 4.2.62). In particular, the FP motif is well located in the groove between helix α1 and α3 in each CUE domain. For the N4BP1 CUE domain, residues from the N-terminus of helix α3 (eg. L884, N885, S888 etc) are likely to be involved in the binding.

However, from the N4BP1^{CUE}/ubiquitin docking model from section 4.2.2.4, the binding interface of N4BP1^{CUE} is restricted to the FP motif and the last residue D893. The other residues with high CSPs are probably affected indirectly by secondary effects upon ubiquitin binding.

4.2.3.2.2 Perturbed surface comparison – from the ubiquitin side

The CSPs from NMR titrations are also mapped to the surface of monoUb and diUb chains respectively. They were also compared with other UBD/ubiquitin complexes.

1) Comparison of the monoUb surfaces

For the perturbed surface of monoUb in the CUE:monoUb pair, the most perturbed area is the I44 hydrophobic patch with the extended surface centred on K48. The importance of K48 for the binding is confirmed by ITC measurements in Figure 4.2.41. There is also another less perturbed area next to the I44 patch which is not directly involved in the interaction from the N4BP1^{CUE}/ubiquitin docking model as shown in section 4.2.2.4.

The perturbed surface centred on the I44 patch and the K48 extension in this study is reminiscent of the binding surface from CUE2-1/ubiquitin complex (PDB:1OTR) (Kang et al., 2003). The key residues involved in the interaction with CUE2-1 are also highlighted on the ubiquitin surface (Figure 4.2.63 right). A surface comparison confirmed that the key residues (V70, L71 etc) from the I44 patch and the K48 extension (G47, K48) are all perturbed upon binding in both cases (Figure 4.2.63). However, despite the similarities, I44 is not involved in the N4BP1^{CUE}/ubiquitin complex, which results in the different binding mode between the two complexes. This will be discussed in section 4.2.3.2.3.

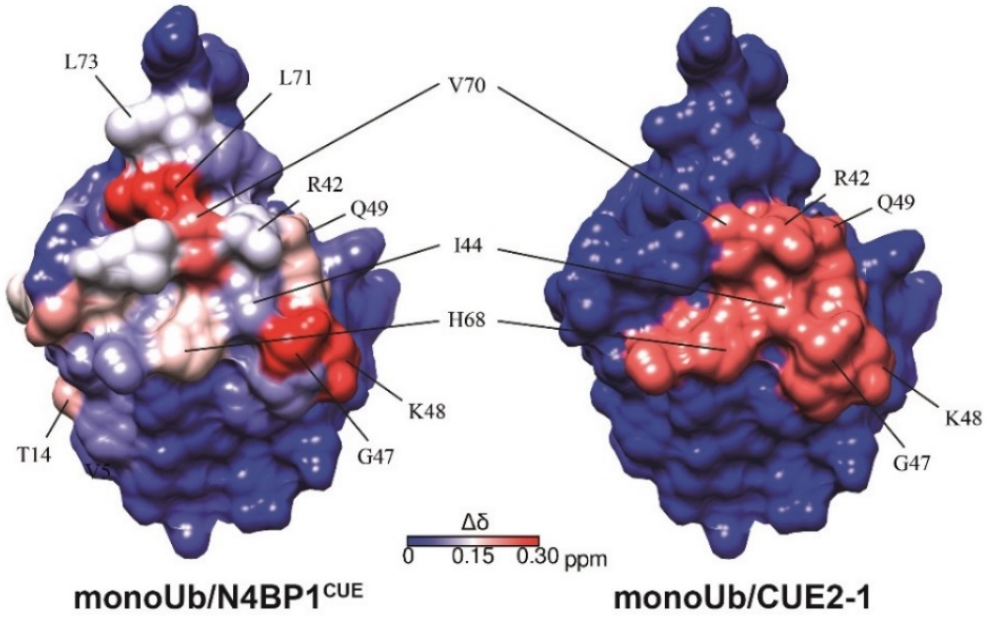


Figure 4.2.63 The comparison of the perturbed surface of monoUb in monoUb/N4BP1^{CUE} and monoUb/CUE2-1. The surface look of monoUb/N4BP1^{CUE} is the same as Figure 4.2.20. For monoUb from monoUb/CUE2-1 complex (PDB: 1OTR), the key residues involved in the binding were coloured in red by CHIMERA according to Kang et al., 2003.

2) Comparison of the M1-diUb and K63-diUb surfaces

Mapping the CSPs to the corresponding surfaces of M1- and K63-diUb indicates that the same binding mode is shared between CUE/M1-diUb, CUE/K63-diUb and CUE/monoUb. Due to the open conformation of both M1- and K63-diUb, one molecule of M1- or K63-diUb could adopt two CUE molecules at the same time. It is also supported by the ITC results from section 4.2.2.1. Thus, the CUE domain does not discriminate between monoUb, M1-diUb and K63-diUb.

The perturbed surfaces of M1- and K63-diUb are further compared with several UBDs which display distinct ubiquitin linkage specificity respectively.

The NEMO CoZi domain has been characterized to specifically bind to linear ubiquitin chains. It has a coiled-coil structure and forms a parallel dimer to interact with M1-diUb (Rahighi et al., 2009). The two ubiquitin

moieties in M1-diUb have divergent binding modes. As shown in Figure4.2.64, the distal ubiquitin contacts the CoZi domain via the canonical I44 patch and the C-terminal tail. The C-terminal tail is part of the M1-linkage region, demonstrating the ability from the CoZi domain to recognize this linkage. In contrast to the distal ubiquitin, the I44 patch from the proximal ubiquitin is not involved in the interaction. Instead, residues that are located next to this I44 patch are employed for the binding. From an overall view, the structure of M1-diUb from this complex appears twisted along the linkage region compared to the crystal structure of the apo M1-diUb, indicating a high degree of rotational flexibility for the ubiquitin chains with M1-linkage.

In comparison, the perturbed surface of M1-diUb upon binding to the N4BP1 CUE domain is different (Figure 4.2.64). First of all, the perturbed surfaces from the two ubiquitin moieties are almost identical in M1-diUb/N4BP1^{CUE}. Key residues from the I44 patch involved in the interactions for both cases are different. Residue L8, I44 and L73 are not strongly perturbed in M1-diUb/N4BP1^{CUE} but crucial to M1-diUb/NEMO^{CoZi}. Besides, the ubiquitin K48 extended surface is employed by M1-diUb/N4BP1^{CUE}, but not in M1-diUb/NEMO^{CoZi}. Although the rotational flexibility of M1-diUb is not considered in the binding study with the N4BP1 CUE domain here, the binding modes between M1-diUb/N4BP1^{CUE} and M1-diUb/NEMO^{CoZi} are still in much difference.

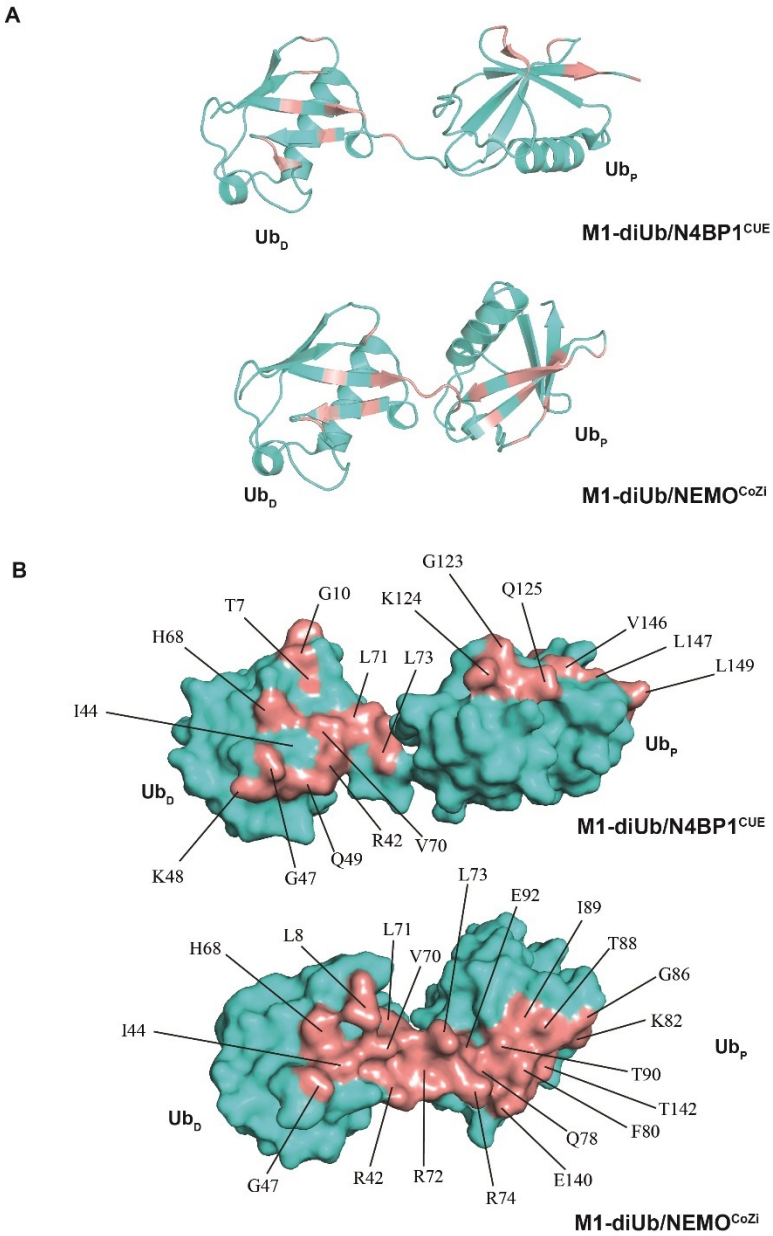


Figure 4.2.64 The comparison between the perturbed surface of M1-diUb from M1-diUb/N4BP1^{CUE} and the binding surface of M1-diUb from M1-diUb/NEMO^{CoZi}. A. The cartoon view of respective M1-diUb chains. B. The surface view of respective M1-diUb chains. The distal ubiquitins from the two structures are placed in the same position for a better comparison. The key perturbed residues (CSP>0.15) on the M1-diUb surface (PDB: 2W9N) upon binding to the N4BP1 CUE domain were highlighted in salmon by Pymol according to the 3X CSP list. The M1-diUb surface from M1-diUb/NEMO^{CoZi} (PDB: 2ZVO) was generated by Pymol and the key residues involved in the binding were highlighted in salmon. Ub$_P$: proximal Ub; Ub$_D$: distal Ub.

As exemplified in section 1.1.3.2, the UIM domains from RAP80 and the NZF domain from TAB2 both show specificity to K63-diUb chains. When it comes to the comparison with K63-diUb/N4BP1^{CUE}, from a general view, the I44 patches from the two ubiquitin moieties in all the three complexes are both involved in the binding (Figure 4.2.65). K63-diUb displays various conformations in complex with different UBDs due to that K63-diUb has a more flexible lysine-linked isopeptide bond, thus, the rotational flexibility of K63-diUb is higher than the M1 linkage.

In the case of the interaction between RAP80^{UIM} and K63-diUb, the proximal ubiquitin interacts with UIM1 while the distal ubiquitin interacts with UIM2. This “one UIM to one Ub” mode seems similar to “one CUE to one Ub”. However, UIM1 and UIM2 are tandemly linked via the inter UIM domain in RAP80 to form a continuous α-helix whereas each N4BP1 molecule only contains one CUE domain. Thus, the binding stoichiometry is different. On the other hand, the K48 extended surface is not involved in K63-diUb/RAP80^{UIM}. Most important of all, the binding specificity of RAP80 comes from the certain length of the inter UIM domain which places the two UIM domains in the most ideal positions to bind to both ubiquitin moieties simultaneously. While on the contrary, the N4BP1 CUE domain does not show binding specificity to K63-diUb.

As for the K63-diUb/TAB^{NZF} complex, one NZF domain binds to both I44 patches in K63-diUb. This specificity highlights the main difference from K63-diUb/N4BP1^{CUE}. Besides, the K48 extended surface is also not involved in K63-diUb/TAB^{NZF}.

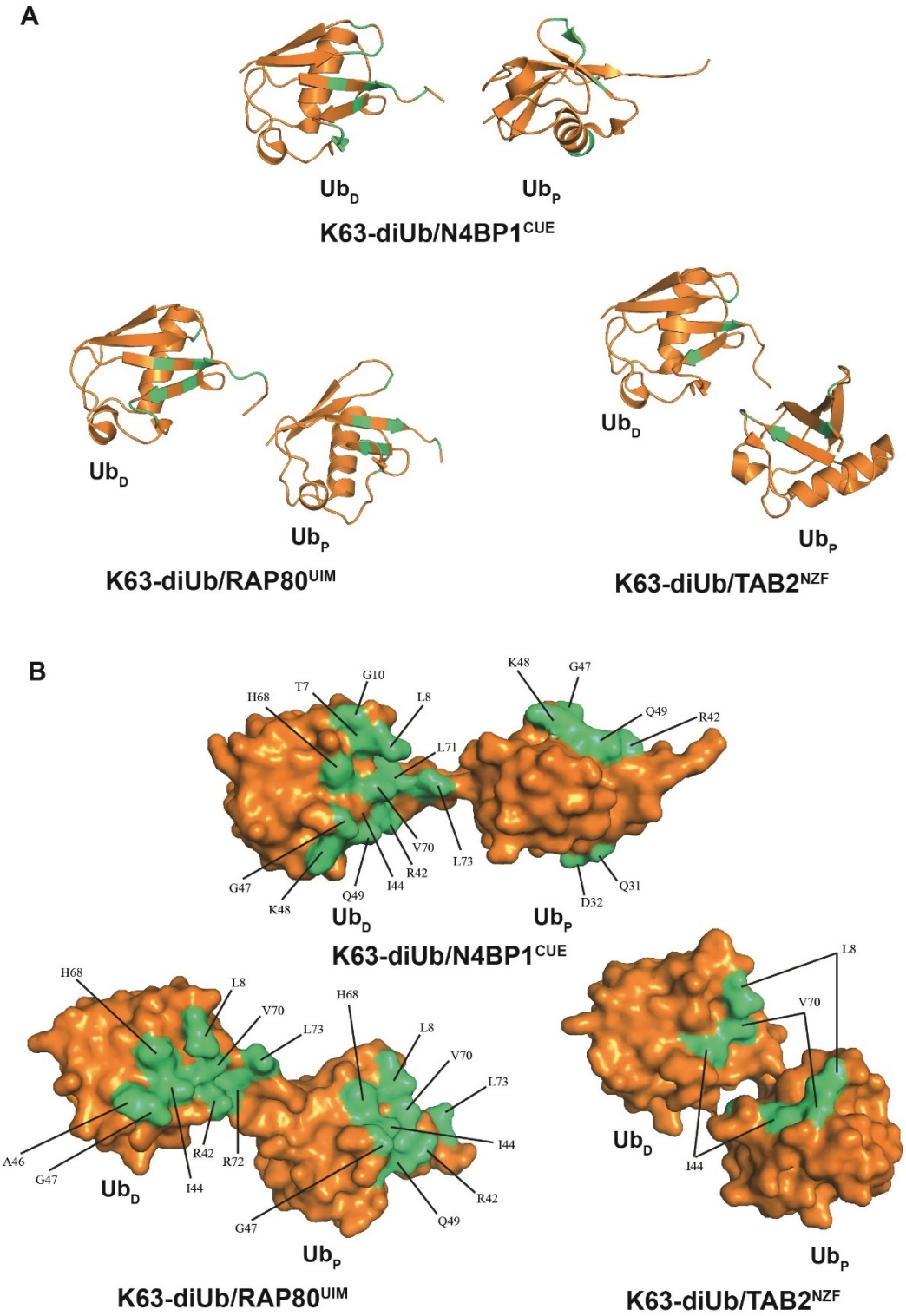


Figure 4.2.65 The comparison between the perturbed surface of K63-diUb from K63-diUb/N4BP1^{CUE} and the binding surface of K63-diUb from K63-diUb/RAP80^{UIM} and K63-diUb/TAB^{NZF} respectively. A. The cartoon view of respective K63-diUb chains. B. The surface view of respective K63-diUb chains. The distal ubiquitins from the three structures are placed in the same position for a better comparison.

The key perturbed residues (CSP>0.15) on the K63-diUb surface (PDB: 3H7P) upon binding to the N4BP1 CUE domain were highlighted in limegreen by Pymol according to the 3X CSP list. The K63-diUb surfaces from K63-diUb/RAP80^{UIM} (PDB: 3A1Q) and K63-diUb/TAB^{NZF} (PDB: 2WWZ) were generated by Pymol respectively. The key residues involved in the binding were highlighted in limegreen. Ub$_P$: proximal Ub; Ub$_D$: distal Ub.

3) Comparison of the K48-diUb surfaces

In section 4.2.2.3.2, the CSP patterns of K48-diUb binding to the CUE domain is similar to the other titration pairs, but with lower intensity. Considering the compact conformation of K48-diUb, it has been demonstrated that K48-diUb has less accessibility to interact with the N4BP1 CUE domain via the canonical monoUb surface.

The binding surface of K48-diUb in the K48-diUb/HHR23A^{UBA} complex is shown in Figure 4.2.66B for the comparison with K48-diUb/N4BP1^{CUE}. In the K48-diUb/HHR23A^{UBA} complex, the I44 patch in the proximal ubiquitin is employed for the binding, which centred on the residue I44. In contrast, the CSPs of I44 in both ubiquitin moieties of K48-diUb are at a much lower level upon binding to the N4BP1 CUE domain. Residue K48 and L73 from the distal ubiquitin are crucial for the binding to HHR23A^{UBA}, however they are less involved in the interaction with the CUE domain.

Although the binding stoichiometries of K48-diUb/HHR23A^{UBA} and K48-diUb/N4BP1^{CUE} both indicate a 1:1 ratio, HHR23A^{UBA} shows the high specificity to recognize K48-diUb, while the N4BP1 CUE domain is not in favour of binding to K48-diUb compared to monoUb, M1-diUb or K63-diUb observed from this study (Table 4.2.6).

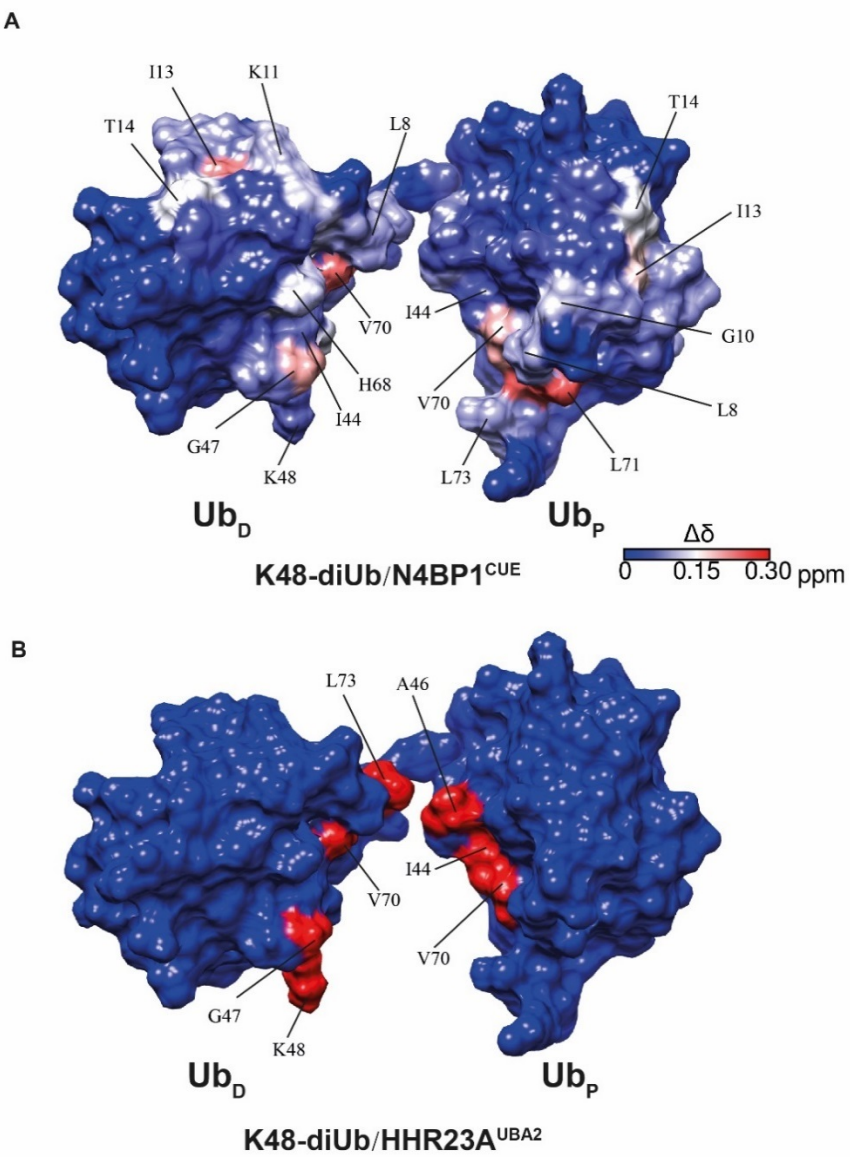


Figure 4.2.66 The comparison between the perturbed surface of K48-diUb from K48-diUb/N4BP1^{CUE} (A) and the binding surface of K48-diUb from K48-diUb/HHR23A^{UBA2} (B). The CSPs of K48-diUb upon binding to the N4BP1 CUE domain were mapped to the K48-diUb structure (PDB:1ZO6) by CHIMERA. The molar ratio between K48-diUb and the CUE domain used here is 3:1 (3X). The K48-diUb surface from K48-diUb/HHR23A^{UBA2} (PDB: 1ZO6) was generated by CHIMERA where the key residues involved in the binding were highlighted in red. Both structures are positioned to the best view to show the ubiquitin I44 patch. Ub$_P$: proximal Ub; Ub$_D$: distal Ub.

4.2.3.2.3 Comparison of the N4BP1^{CUE}/ubiquitin complex with other complexes

In section 4.2.2.4, based on the CSPs from NMR titration, the complex model between the N4BP1 CUE domain and monoUb was generated. The complex model highlights the hydrophobic interaction between the CUE FP motif and ubiquitin I44 patch, as well as the polar interaction between CUE D893 and ubiquitin K48.

This complex model is also compared with another two CUE/ubiquitin complexes (Figure 4.2.67). From an overall view, the orientation of the N4BP1 CUE domain is rotated ~ 90° compared to the CUE domains positioned in the other two complexes. Although the ubiquitin I44 hydrophobic patch is involved in the interface in all the three complexes, the CUE domains from both CUE2-1 and gp78 directly contact residue I44 from ubiquitin while the different CUE orientation in the N4BP1^{CUE}/ubiquitin complex does not employ I44 for the interaction.

It is worth noticing that ubiquitin K48 binds to both CUE domains via polar interaction in the N4BP1^{CUE}/ubiquitin complex and the CUE2-1/ubiquitin complex. Structural study reveals that residue D51 from CUE2-1 interacts with ubiquitin K48 in the complex. It was proposed that CUE2-1 may play an inhibitory role in the formation of K48-linked ubiquitin chains (Kang et al., 2003). On the contrary, ubiquitin K48 is not involved in the gp78^{CUE}/ubiquitin complex.

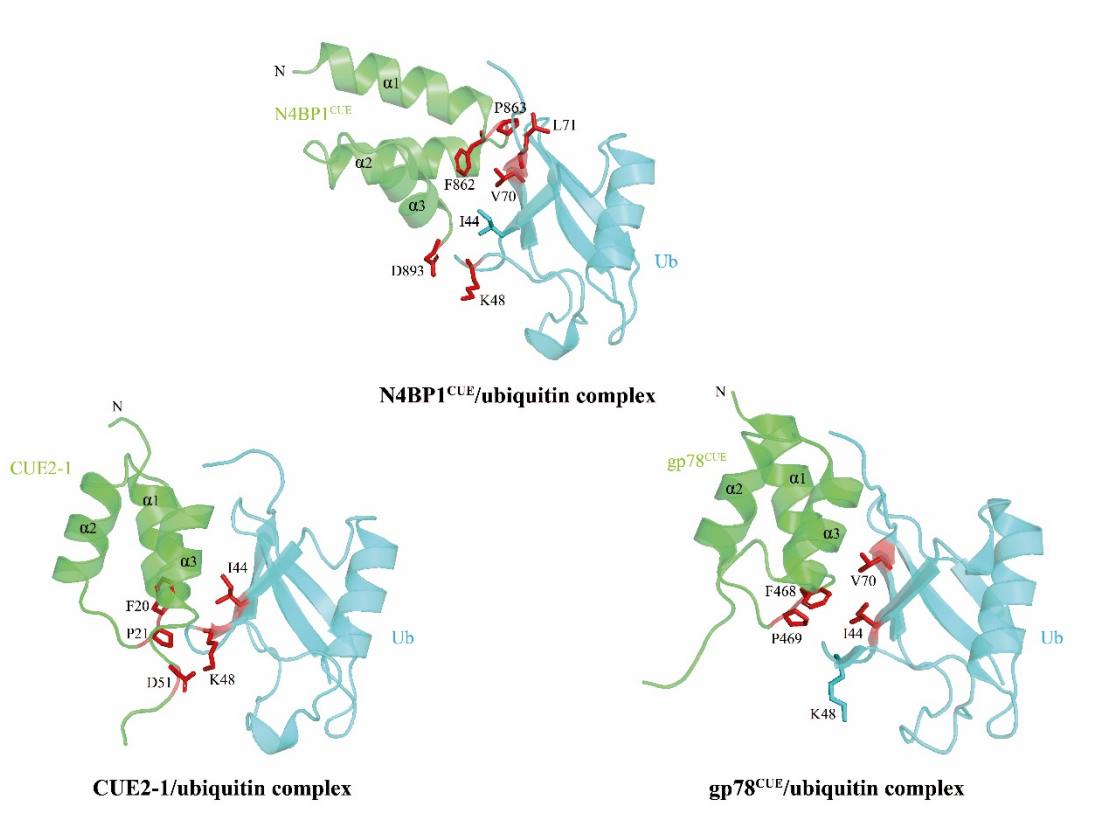


Figure 4.2.67 Comparison of three CUE domains in complex with monoUb. Ubiquitin from each complex is placed in the same position for a better comparison. Key residues involved in the interaction are coloured in red. The CUE domain is coloured in green while ubiquitin is coloured in cyan. Figures were generated by Pymol using the docking model for the N4BP1^{CUE}/ubiquitin complex, PDB: 1OTR (CUE2-1/ubiquitin complex) and PDB: 2LVO (gp78^{CUE}/ubiquitin complex).

The model of the N4BP1 CUE domain in complex with monoUb is also superimposed onto the structures of M1-diUb, K63-diUb and K48-diUb respectively. As shown in Figure 4.2.68, two molecules of CUE could simultaneously bind to the two ubiquitin moieties from either M1-diUb or K63-diUb. This is also in accordance with the 2:1 ratio of CUE:M1-diUb or CUE:K63-diUb observed from the ITC studies. However, in the case of K48-diUb, the steric obstruction allows only one CUE molecule to be adopted. As suggested in Figure 4.2.68, if the N4BP1^{CUE}/ubiquitin model was superimposed onto the proximal ubiquitin of K48-diUb, the key residues K48,

V70 and L71 from the distal ubiquitin are facing to the CUE domain. Therefore, there is not enough space to adopt another CUE molecule. This model also supports the 1:1 ratio of CUE:K48-diUb measured by ITC.

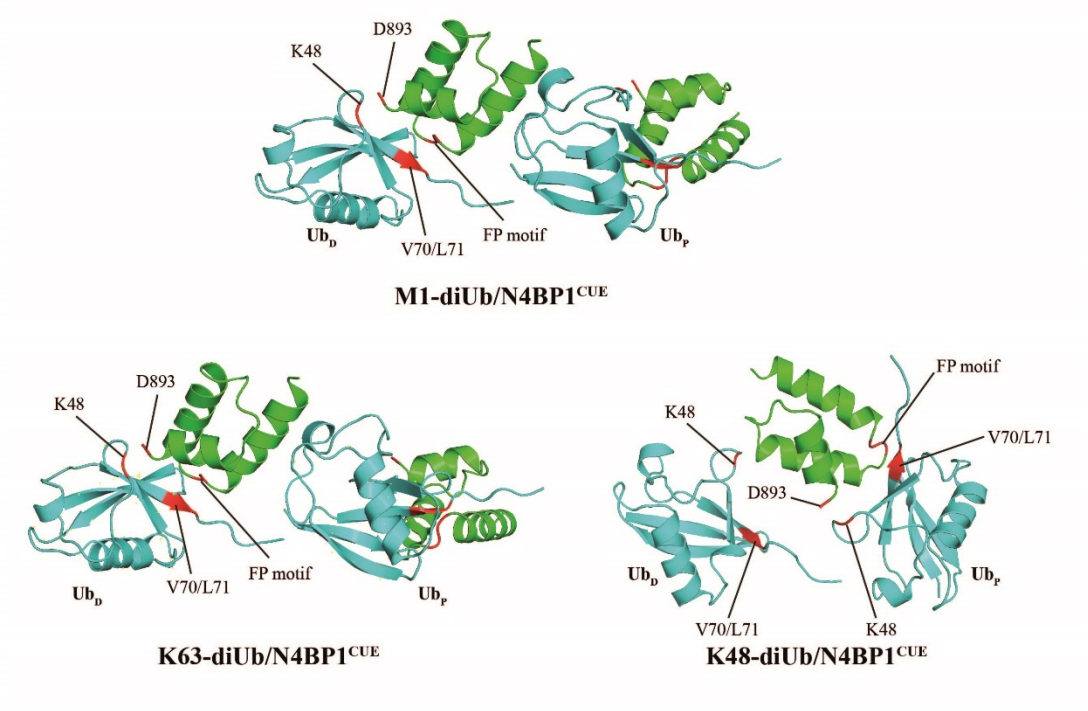


Figure 4.2.68 Structure models of the N4BP1 CUE domain interacting with different types of di-ubiquitin chains. Key residues involved in the interaction are indicated in red. The N4BP1 CUE domain is coloured in green while di-ubiquitin chains are coloured in cyan. Ub$_P$: proximal Ub; Ub$_D$: distal Ub. Figures were generated by superimposing the N4BP1^{CUE}/ubiquitin complex model to M1-diUb (PDB: 2W9N), K63-diUb (PDB: 3H7P) and K48-diUb (PDB:1ZO6) respectively.

As mentioned in section 4.2.3.1, the structure of human N4BP1 CUE domain (the CoCUN domain) has been solved by solution NMR by another research group. The CoCUN/ubiquitin complex structure is also present from the same study (Nepravishta et al., 2019), but again absent in the PDB database. According to their descriptions, several conflicts were found in comparison with the N4BP1^{CUE}/ubiquitin complex model generated from this thesis.

First of all, ubiquitin I44 lacks a measurable CSP from the CoCUN/ubiquitin complex, which is in line with the observation from this work. However, this complex still shows that ubiquitin I44 is involved in the interaction with CoCUN based on the observation where ubiquitin I44A mutant abolished the interaction.

In contrast to the structural findings of this work, the valine and leucine residues at the end of helix α3 in CoCUN did not show any CSPs. The C-terminal tail of CoCUN is not involved in the interaction according to the CoCUN/ubiquitin complex, while in the N4BP1^{CUE}/ubiquitin complex presented in this thesis, the C-terminal tail of N4BP1^{CUE} forms polar interaction with residue K48 from ubiquitin. Strong CSPs of the key residues were observed from both N4BP1^{CUE} and ubiquitin. This is also supported by the ITC data using the K48 mutants.

In addition, ubiquitin K48 also has strong perturbation according to the CSP data for the interaction between CoCUN and ubiquitin. However, Nepravishta and colleagues did not consider or include it in their CoCUN/ubiquitin complex. Due to that the structure of CoCUN/ubiquitin is not deposited in the PDB database, superimposing this complex to different diUbs cannot be achieved. Nevertheless, the CoCUN/ubiquitin complex could not explain the ability that N4BP1^{CUE} distinguishes K48-diUb from other types of ubiquitin.

In all, considering the inaccuracy of the CoCUN structure, it is also questionable that the CoCUN/ubiquitin complex is reliable.

4.2.3.3 Structural model of dimer-induced specificity of N4BP1 for linear ubiquitin chains

In section 4.2.2.1 and 4.2.2.3, the ubiquitin-binding properties of the N4BP1 CUE domain were investigated while in section 4.2.2.5, the ubiquitin-binding properties of the N4BP1 construct containing both the RNAse domain and the CUE domain were studied. A summary of the Kd values collected from this while chapter is shown in Figure 4.2.69.

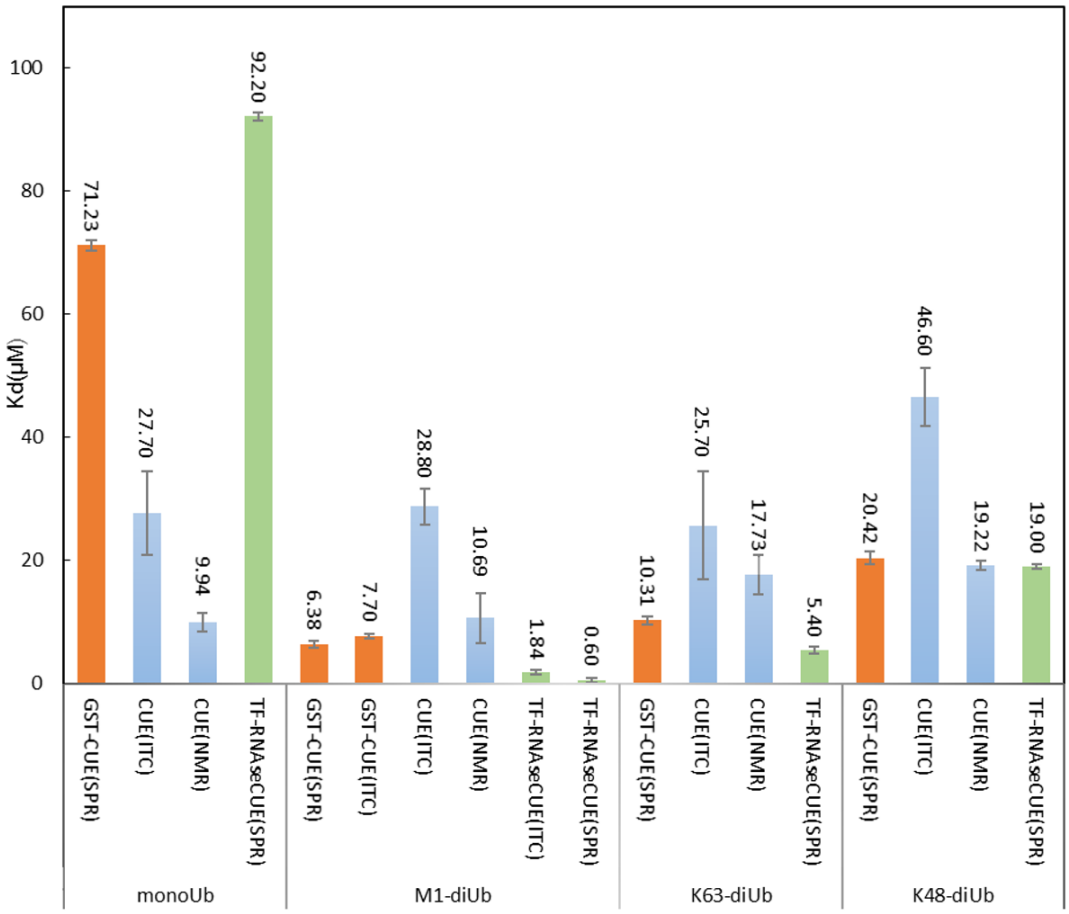


Figure 4.2.69 The summary of the Kd values from this study. The Kd value is indicated above each bar. The data are colour-coded according to the protein used for the interaction (The data of GST-CUE are coloured in orange while CUE in sky blue and TF-RNAseCUE in green).

The N4BP1 CUE domain could only distinguish K48-diUb from monoUb, M1-diUb and K63-diUb. The affinity of the CUE domain to monoUb, M1-diUb and K63-diUb are comparable. The RNAseCUE construct (including both the RNAse domain and the CUE domain) displays higher affinity to M1-diUb compared to monoUb, K48-diUb and K63-diUb. The tight binding between RNAseCUE and M1-diUb is further confirmed by ITC. However, the N4BP1 RNAse domain does not interact with ubiquitin under the experimental condition used here, which makes the CUE domain as the sole ubiquitin-binding module in the N4BP1 C-terminus. This suggests that the RNAse domain has an indirect role in mediating ubiquitin binding via the CUE domain. Further investigation by analytical sec-exclusion chromatography demonstrated that the RNAse domain forms oligomers (including dimer and tetramer).

Based on the above experimental evidence, a structural model of dimer-induced specificity of N4BP1 for linear ubiquitin chains is proposed in this section.

The structure prediction of the whole C-terminal N4BP1 construct RNAseCUE was performed by trRosetta. The secondary structure of the linker region (775-849) between the RNAse domain and the CUE domain was predicted by Phyre2. As seen from Figure 4.2.70, the trRosetta prediction provided five top models and the straight-line length of this linker is within a range of 40-60 Å. The linker between the RNAse domain and the CUE domain is only partially structured and indicates high flexibility. Phyre2 could not find the structure template with confidence > 5% for this linker. Superimposing the RNAse domains from all the 5 models together indicates that the movement of the CUE domain is very flexible within a large range of space due to this linker.

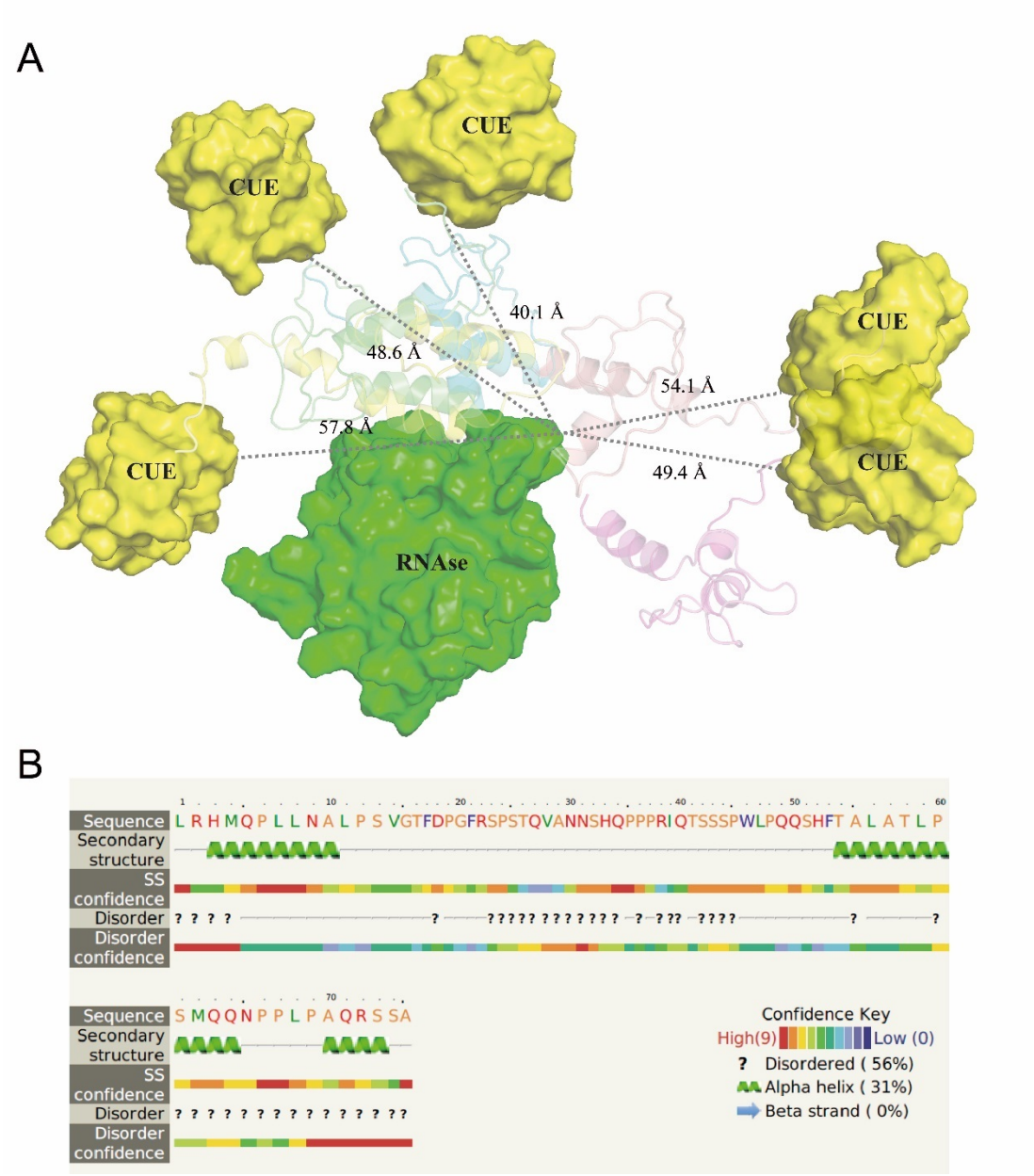


Figure 4.2.70 Superimposition of the five top models of the N4BP1 C-terminus (RNAseCUE) (A) and secondary structure prediction of the linker region (B). The models were predicted by trRosetta. The RNAse domains from the five models were superimposed and coloured in green. The five CUE domains are coloured in yellow. The RNAse domain and the CUE domain are shown in the surface mode, while the five flexible linkers are shown in cartoon. The grey dashed lines indicate the straight-line distance (ranging from 40 to 60 Å) between the RNAse domain and the CUE domain. The secondary structure prediction of the linker region (775-849) was performed by Phyre2. Figure generated by Pymol.

The predicted CUE domain is highly similar to the CUE structure model described in section 4.2.1.2.5 (RMSD=0.895, see the discussion in section 4.2.3.1), which allows the superimposition of the N4BP1^{CUE}/ubiquitin docking model to the predicted RNAseCUE structure (The top model 1 of RNAseCUE was used here, Figure 4.2.71A). Furthermore, as mentioned in section 4.2.2.5.4, the oligomerization of the RNAse domain was proposed to be able to bring several CUE domains in close distance. Since the crystal structure of the MCPIP1 RNAse head-to-tail dimer is available (PDB: 5H9V), an N4BP1 RNAseCUE head-to-tail dimer model was generated based on this crystal structure (Figure 4.2.71B). Then the M1-diUb/N4BP1^{CUE} model from Figure 4.2.68 was superimposed onto the CUE domain of this dimer model. The top model 1 of RNAseCUE was used as one of the protomers. For another protomer, the distance between the CUE N-terminus and the RNAse C-terminus was estimated at ~ 60 Å by Pymol. This could be possibly achieved by one of the models as presented in Figure 4.2.70.

The above model only considers the RNAse dimer for the binding. However, the RNAse or the RNAseCUE also displays tetrameric states as demonstrated in section 4.2.2.5.4. In theory, this effect could increase the chance for the N4BP1 CUE domain binding to longer linear ubiquitin chains. This has been experimentally proved by the ITC study in section 4.2.2.5.4.

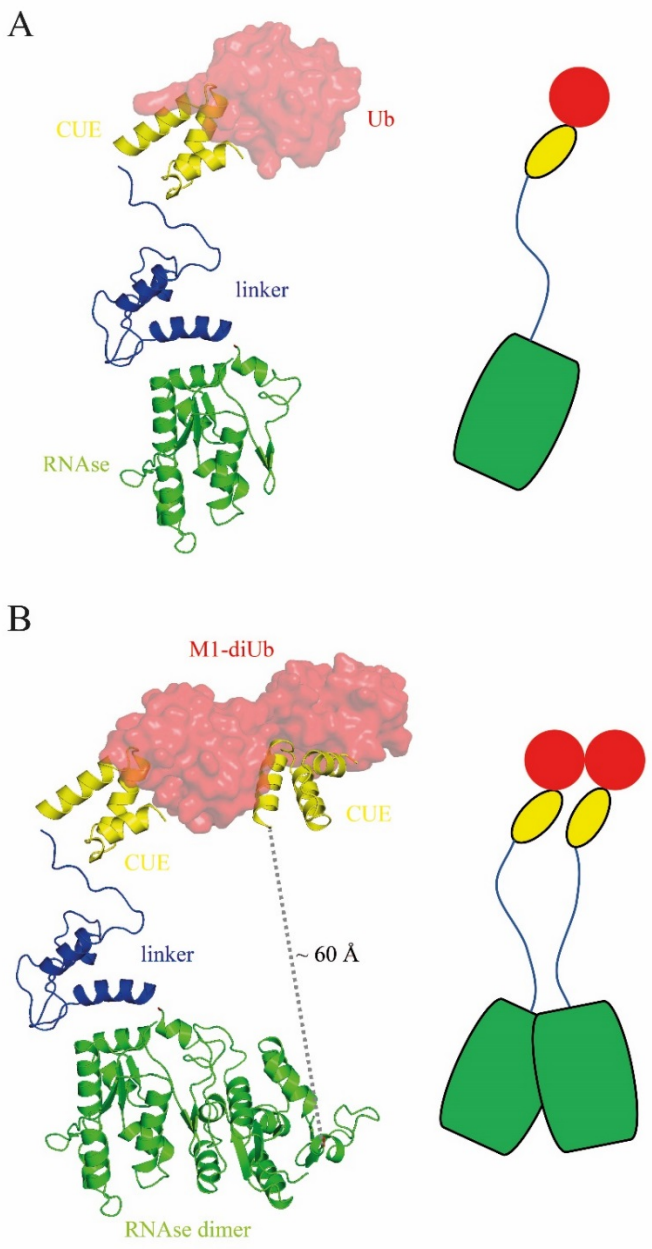


Figure 4.2.71 Working models of the N4BP1 C-terminal construct (RNAseCUE) interacting with ubiquitin. A. Superimposition of the top model 1 of the N4BP1 C-terminus with the N4BP1^{CUE}/ubiquitin docking model. B. The RNAseCUE dimer model. The top model 1 of the N4BP1 C-terminus was used as one of the protomers on the left. The grey dashed lines indicate the straight-line distance between the RNAse domain and the CUE domain in another RNAseCUE protomer. MonoUb and M1-diUb are shown in the surface mode. The schematic representation of the structure model is shown on the right of each figure. The top model 1 of the N4BP1 C-terminus was predicted by trRosetta. Figure generated by Pymol.

In this section, a working model of the interaction between the dimeric C-terminal N4BP1 and linear ubiquitin chains was discussed. This model describes a novel ubiquitin-binding mode which helps explain the high affinity of the interaction from two aspects (Figure 4.2.72):

1) As a general UBD, the CUE domain binds to ubiquitin in a moderate affinity. It could only discriminate K48-linked ubiquitin chains from monoUb, M1- and K63-linked chains;

2) The RNAse domains could form head-to-tail oligomers to increase the local concentration of the C-terminal CUE domains and possibly re-organise the CUE domains in an ideal spatial arrangement, hence increasing the affinity to interact with the longer linear ubiquitin chains.

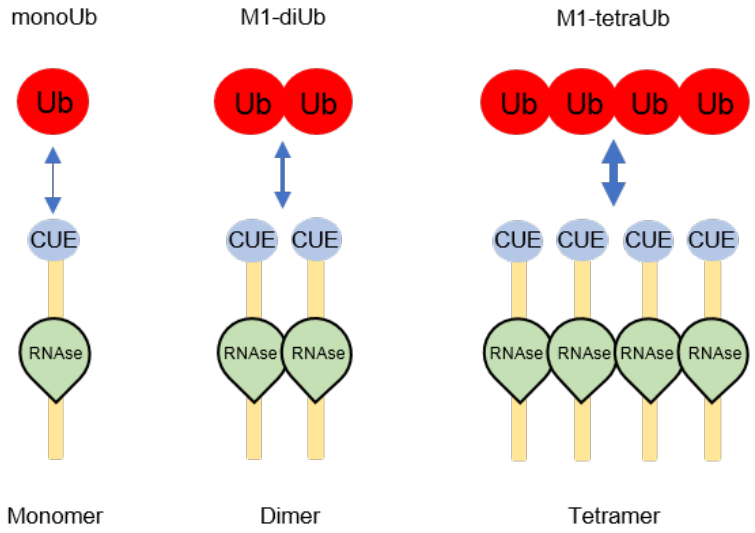


Figure 4.2.72 Schematic representation of the monomeric, dimeric and tetrameric N4BP1 RNAseCUE binding to linear ubiquitin chains.

4.2.3.4 N4BP1/MCPIP1/KHNYN comparison

As introduced in Chapter I, the RNAse domains of N4BP1, KHNYN and MCPIP family member MCPIP3 are grouped close together based on a phylogenetic study (Anantharaman and Aravind, 2006; Figure 1.27).

In Chapter 4.1, it has been shown that the top template of the predicted N4BP1 UBA-like domain by Phyre2 was the NTD domain from MCPIP1 (Figure 4.1.8 and 4.1.9). From the study in Chapter 4.2, the structure comparison by DALI server found that the N4BP1 CUE domain is similar to the NCPIP1 CTD domain (Figure 4.2.57). In addition, although the N4BP1 CUE domain and the KHNYN CUBAN domain share high sequence similarity by a Phyre2 search, the structures of the two domains are proved to be different (Figure 4.2.61). In addition, N4BP1 RNAse domain shares high sequence similarity to the MCPIP1 RNAse domain (Figure 4.2.45). The RNAse activity from both domains has been experimentally confirmed (Matsushita et al., 2009; Xu et al., 2012; Yamasoba et al., 2019; Yokogawa et al., 2016).

Therefore, a careful domain comparison between the 3 proteins was conducted and displayed in Figure 4.2.73. The full-length sequence alignments are summarized in Appendix III. Interestingly, apart from the high similarity between N4BP1 and MCPIP1 from the above findings, the sequence alignments also indicate that N4BP1 and KHNYN are very similar to each other except that the UBA-like domain is absent in KHNYN. The RNAse domain of KHNYN does not have the catalytic Cys residue for DUB activity.

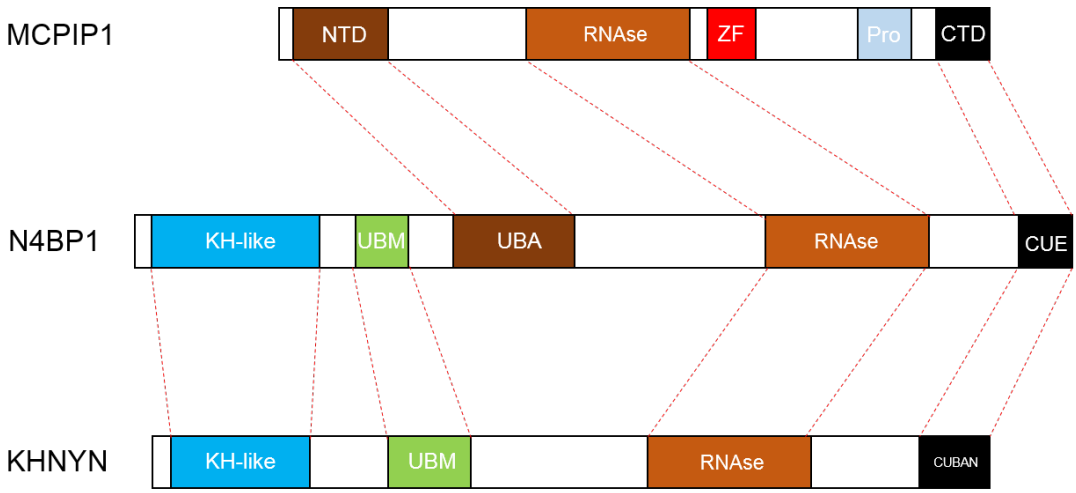


Figure 4.2.73 Domain comparison between MCPIP1, N4BP1 and KHNYN. CTD: C-terminal domain. CUBAN: Cullin-binding domain associating with NEDD8. KH: K Homology domain. NTD: N-terminal domain. Pro: Proline-rich domain. UBA: Ubiquitin association domain. UBM: Ubiquitin binding motif. ZF: CCCH-type zinc-finger motif.

There are several functional studies of MCPIP1 as summarized in section 1.4.3, which revealed the negative roles of MCPIP1 in regulating NFκB activation. The study of KHNYN is limited. The interaction between KHNYN CUBAN and NEDD8 suggests a role of KHNYN in regulating the E3 ligase activity of CRLs. The functional study of N4BP1 is mainly centred on its RNAse activity in degrading virus RNA. Until recently, N4BP1 is identified as a novel linear ubiquitin receptor and negatively regulates the TNFR signaling pathway from the cell biological studies by the research group of Koraljka Husnjak.

These sequence and structure similarities between the three proteins could possibly offer some new perspectives to study the function of each protein. For example, the structure of each domain from MCPIP1 is available, which is helpful for the structure investigation of N4BP1 and KHNYN. The function of MCPIP1 and N4BP1 in negatively regulating inflammatory signaling could also possibly provide some new directions to study the function of KHNYN.

CHAPTER V

Summary and Conclusions

In this thesis, two functions of N4BP1, inhibition of LUBAC and ubiquitin binding have been investigated.

In Chapter 4.1, the N-terminus of N4BP1 was identified as the inhibitory module. Based on the observation from *in vitro* ubiquitin assays, the UBA-like domain strongly inhibits the E3 ligase activity of HOIP, while the KH-like domain displays a much weaker inhibitory effect. Pull-down assays have demonstrated that the HOIP binding region from N4BP1 covers the UBM-like domain and it is important to modulate the inhibition. In all, the three domains from the N4BP1 N-terminus cooperate as a single functional module to achieve optimal inhibition of HOIP.

Chapter 4.2 are focused on the linear ubiquitin-binding specificity of N4BP1. The NMR structure of the CUE domain was solved in section 4.2.2.2. Although sequence analysis implies that the N4BP1 CUE domain shares high sequence similarity with the CUBAN domain from KHNYN, however, the CUE domain adopts a conserved protein fold shared with other canonical CUE domains. The FP motif from the N4BP1 CUE domain is also conserved among the CUE domain family and plays a key role in the interaction with ubiquitin.

In section 4.2.2.1, the ubiquitin-binding properties of the CUE domain were quantitively characterized by SPR and ITC. SPR measurements have shown that the GST-tagged CUE domain prefers binding to M1-diUb compared to monoUb, K63- and K48-diUb. However, the ITC measurements using the tag-free CUE domain demonstrated that the CUE domain only has a lower affinity to K48-diUb, while its affinity to monoUb, M1-diUb and K63-diUb is similar. Additional SEC-MALLS analysis indicated that the GST-tagged CUE domain is in the dimer form due to the dimerization of the GST tag. The GST dimer brings the two CUE molecules in close proximity, thus artificially increasing the affinity to M1-diUb.

To identify the perturbed surfaces from both the CUE domain and ubiquitin in the interaction, a series of NMR titrations were performed in

section 4.2.2.3. The main perturbed surface of the CUE domain consists of a groove between helix α1 and α3, where the conserved FP motif is located. From an overall view, the CSP patterns of the CUE domain in different CUE:ubiquitin pairs are similar, indicating that the CUE domain deploys a conserved binding mode to interact with monoUb or diUb chains. The CSP patterns of different ubiquitin chains are also correlated with the corresponding surfaces. The perturbed surface of monoUb upon binding to the CUE domain highlights the key involvement of the I44 hydrophobic patch and the K48 extended surface in the interaction. The distal and proximal ubiquitin moieties from M1- and K63-diUb also have similar CSP patterns to monoUb. As M1- and K63-diUb both adopt an open conformation, it allows the I44 patch as well as the K48 extension from both proximal and distal ubiquitin moieties well exposed to the solvent. Therefore, two CUE domain molecules form the same independent interface with the two ubiquitin moieties as observed for the monoUb. In the case of K48-diUb, the I44 patch and the K48 extended surface from both proximal and distal ubiquitin were much less perturbed, which is in line with the lower affinity observed between CUE and K48-diUb by ITC. Notably, one of the K48 residues from the two ubiquitin moieties in K48-diUb is occupied to form the K48 linkage, thus the K48 extended surface in the proximal ubiquitin is not accessible to the CUE domain, suggesting that only one CUE molecule could interact with one K48-diUb molecule. This is in agreement with the 1:1 stoichiometry between the CUE domain and K48-diUb observed from ITC. In all, the study in section 4.2.2.3 demonstrates that the CUE domain discriminates against K48-diUb interaction, while binding to M1-, K63- and monoUb is in a nonselective fashion.

In section 4.2.2.4, a docking model for the interaction between the CUE domain and monoUb was presented. The FP motif of the CUE domain interacts with part of the ubiquitin I44 patch (V70 and L71) while I44 is not involved. Residue D893 in the C-terminal tail of the CUE domain forms polar

interaction with residue K48 from ubiquitin. These two binding properties place the CUE molecule onto the ubiquitin surface in a different orientation compared to the other canonical CUE/ubiquitin complexes. Superimposing this model to the structure of M1- or K63-diUb indicates that two CUE molecules could well bind to the two ubiquitin moieties at the same time. While on the contrary, K48-diUb could only adopt one CUE molecule following this model. Again, this is in agreement with the stoichiometry from the experimental evidence by ITC.

In section 4.2.2.5, the involvement of the RNAse domain in the ubiquitin binding was investigated. The N4BP1 C-terminal construct RNAseCUE (containing both the RNAse domain and the CUE domain) displays high affinity to M1-diUb by both SPR and ITC studies. Interestingly, the RNAse domain doesn't interact with ubiquitin, suggesting its indirect role in mediating ubiquitin binding while the CUE domain is the sole ubiquitin-binding domain in the N4BP1 C-terminus. Similar to its homologous RNAse domain in MCPIP1, the analytical size-exclusion chromatography demonstrated that the N4BP1 RNAse domain also forms oligomers (including dimers and tetramers). Based on the evidence, a structural model was proposed in section 4.2.3.3 to illustrate the dimer-induced specificity of N4BP1 for linear ubiquitin chains. In this model, the dimerization of the RNAse domain brings two CUE molecules in close vicinity, thus the local concentration of the N4BP1 C-terminus is increased, which facilitates the binding to linear diUb chains. This suggests that oligomerization of the RNAse domain could induce appropriate structural arrangements of N4BP1 to facilitate recognition for linear ubiquitin chains. This is further supported by the ITC study where the RNAseCUE construct displays higher affinity to M1-tetraUb than M1-diUb.

Taken together, the two modules from N4BP1 facilitate its negative regulation of LUBAC activity in TNFR signaling. On one hand, the N-terminal N4BP1 inhibits the E3 ligase activity of LUBAC, which results in the reduced

synthesis of linear ubiquitin chains (Figure 5.1 ①). On the other hand, the C-terminal N4BP1 might competitively bind to linear ubiquitin chains, thereby prevents the linear ubiquitination to the substrates catalyzed by LUBAC (Figure 5.1 ②). In addition, the N4BP1 C-terminus could also compete with other key modulators (eg. NEMO) in the TNFR signaling for binding to linear ubiquitin chains, thus disrupting the downstream substrate recruitments in this signaling pathway (Figure 5.1 ③). All these effects impair the TNFR signaling mediated by linear ubiquitin chains.

The study in this thesis provides a wider understanding of the negative regulatory roles of N4BP1 in TNFR signaling from a biochemical and structural perspectives. It will be interesting to investigate how the three domains from N4BP1 N-terminus cooperate to inhibit HOIP. Structural approaches which elucidate the complex of N4BP1 with HOIP will ultimately shed light on the molecular basis for the function of N4BP1.

Furthermore, it will be important to solve the structure of the dimeric or oligomeric form of N4BP1 in complex with linear ubiquitin chains to confirm the model proposed in this thesis. The N4BP1 RNAse domain is homologous to the RNAse domain in MCPIP1. Although the crystal structure of the dimeric MCPIP1 RNAse domain has been solved, it is challenging to express the N4BP1 RNAse domain for structural studies using the prokaryotic expression system. Therefore, the expression of this domain requires appropriate optimization in the future. Eukaryotic expression systems using mammalian or insect cells might be helpful to produce the protein with high quality.

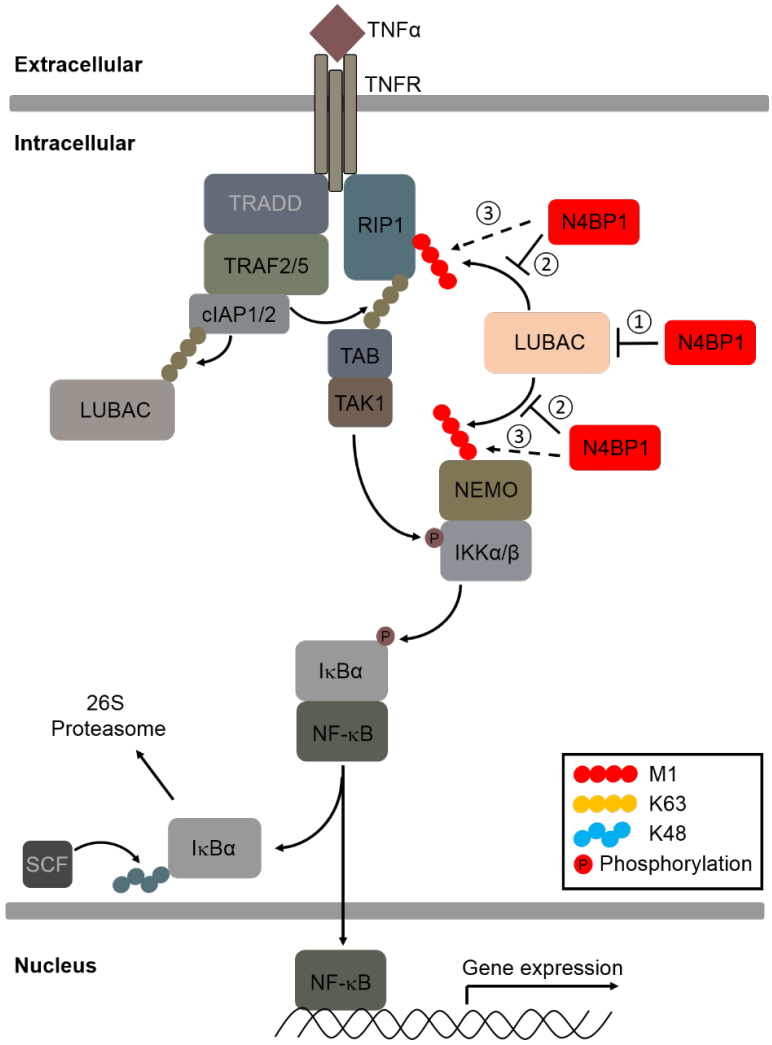


Figure 5.1 Schematic representation of the modulation by N4BP1 in the TNFR signaling. Key components regulated by N4BP1 are highlighted while others are shaded. ①: The inhibition of the E3 ligase activity of LUBAC by the N4BP1 N-terminus. ②: The N4BP1 C-terminus prevents the linear ubiquitination to the substrates catalyzed by LUBAC. ③: The N4BP1 C-terminus competes with key modulators such as NEMO in the TNFR signaling for binding to linear ubiquitin chains.

References

Aguilera, M., Oliveros, M., Martinez-Padron, M., Barbas, J. A., & Ferrus, A. (2000). Ariadne-1: a vital Drosophila gene is required in development and defines a new conserved family of ring-finger proteins. *GENETICS*, 155(3), 1231-1244.

Alam, S. L., Sun, J., Payne, M., Welch, B. D., Blake, B. K., Davis, D. R., Meyer, H. H., Emr, S. D., & Sundquist, W. I. (2004). Ubiquitin interactions of NZF zinc fingers. *EMBO JOURNAL*, 23(7), 1411-1421.

Al-Hakim, A. K., Zagorska, A., Chapman, L., Deak, M., Peggie, M., & Alessi, D. R. (2008). Control of AMPK-related kinases by USP9X and atypical Lys29/Lys33-linked polyubiquitin chains. *BIOCHEMICAL JOURNAL*, 411(2), 249-260.

Anantharaman, V., & Aravind, L. (2006). The NYN domains: novel predicted RNAses with a PIN domain-like fold. *RNA Biology*, 3(1), 18-27.

Barnhart, B. C., & Peter, M. E. (2003). The TNF receptor 1: a split personality complex. *CELL*, 114(2), 148-150.

Bennett, E. J., & Harper, J. W. (2008). DNA damage: ubiquitin marks the spot. *NATURE STRUCTURAL & MOLECULAR BIOLOGY*, 15(1), 20-22.

Boesch, C. (2004). Nobel Prizes for nuclear magnetic resonance imaging: 2003 and historical perspectives. *JOURNAL OF MAGNETIC RESONANCE IMAGING*, 19(5), 517-519.

Brazee, P., Dada, L. A., & Sznajder, J. I. (2016). Role of Linear Ubiquitination in Health and Disease. *Am J Respir Cell Mol Biol*, 54(6), 761-768.

Castagnoli, L., Mandaliti, W., Nepravishta, R., Valentini, E., Mattioni, A., Procopio, R., Iannuccelli, M., Polo, S., Paci, M., Cesareni, G., & Santonico, E. (2019). Selectivity of the CUBAN domain in the recognition of ubiquitin and NEDD8. *The FEBS Journal*, 286(4), 653-677.

Chastagner, P., Israel, A., & Brou, C. (2006). Itch/AIP4 mediates Deltex degradation through the formation of K29-linked polyubiquitin chains. *EMBO REPORTS*, 7(11), 1147-1153.

Chen, L., & Madura, K. (2002). Rad23 promotes the targeting of proteolytic substrates to the proteasome. *MOLECULAR AND CELLULAR BIOLOGY*, 22(13), 4902-4913.

Chen, Z. J., & Sun, L. J. (2009). Nonproteolytic functions of ubiquitin in cell signaling. *MOLECULAR CELL*, 33(3), 275-286.

Chung, K. K., Zhang, Y., Lim, K. L., Tanaka, Y., Huang, H., Gao, J., Ross, C. A., Dawson, V. L., & Dawson, T. M. (2001). Parkin ubiquitinates the alpha-synuclein-interacting protein, synphilin-1: implications for Lewy-body formation in Parkinson disease. *NATURE MEDICINE*, 7(10), 1144-1150.

Ciechanover, A., & Schwartz, A. L. (1998). The ubiquitin-proteasome pathway: the complexity and myriad functions of proteins death. *Proc Natl Acad Sci U S A*, 95(6), 2727-2730.

Cunningham, C. N., Baughman, J. M., Phu, L., Tea, J. S., Yu, C., Coons, M., Kirkpatrick, D. S., Bingol, B., & Corn, J. E. (2015). USP30 and parkin homeostatically regulate atypical ubiquitin chains on mitochondria. *NATURE CELL BIOLOGY*, 17(2), 160-169.

Da, C. C., Sunyach, C., Giaime, E., West, A., Corti, O., Brice, A., Safe, S., Abou-Sleiman, P. M., Wood, N. W., Takahashi, H., Goldberg, M. S., Shen, J., & Checler, F. (2009). Transcriptional repression of p53 by parkin and impairment by mutations associated with autosomal recessive juvenile Parkinson's disease. *NATURE CELL BIOLOGY*, 11(11), 1370-1375.

Damgaard, R. B., Fiil, B. K., Speckmann, C., Yabal, M., Zur, S. U., Bekker-Jensen, S., Jost, P. J., Ehl, S., Mailand, N., & Gyrd-Hansen, M. (2013). Disease-causing mutations in the XIAP BIR2 domain impair NOD2-dependent immune signalling. *EMBO Molecular Medicine*, 5(8), 1278-1295.

Damgaard, R. B., Nachbur, U., Yabal, M., Wong, W. W., Fiil, B. K., Kastirr, M., Rieser, E., Rickard, J. A., Bankovacki, A., Peschel, C., Ruland, J., Bekker-Jensen, S., Mailand, N., Kaufmann, T., Strasser, A., Walczak, H., Silke, J., Jost, P. J., & Gyrd-Hansen, M. (2012). The ubiquitin ligase XIAP recruits LUBAC for NOD2 signaling in inflammation and innate immunity. *MOLECULAR CELL*, 46(6), 746-758.

Dove, K. K., & Klevit, R. E. (2017). RING-Between-RING E3 Ligases: Emerging Themes amid the Variations. *JOURNAL OF MOLECULAR BIOLOGY*, 429(22), 3363-3375.

Duda, D. M., Olszewski, J. L., Schuermann, J. P., Kurinov, I., Miller, D. J., Nourse, A., Alpi, A. F., & Schulman, B. A. (2013). Structure of HHARI, a RING-IBR-RING ubiquitin ligase: autoinhibition of an Ariadne-family E3 and insights into ligation mechanism. *STRUCTURE*, 21(6), 1030-1041.

Dynek, J. N., Goncharov, T., Dueber, E. C., Fedorova, A. V., Izrael-Tomasevic, A., Phu, L., Helgason, E., Fairbrother, W. J., Deshayes, K., Kirkpatrick, D. S., & Vucic, D. (2010). c-IAP1 and UbcH5 promote K11-linked polyubiquitination of RIP1 in TNF signalling. *EMBO JOURNAL*, 29(24), 4198-4209.

Eiyama, A., & Okamoto, K. (2015). PINK1/Parkin-mediated mitophagy in mammalian cells. *CURRENT OPINION IN CELL BIOLOGY*, 33, 95-101.

Elmehdawi, F., Wheway, G., Szymanska, K., Adams, M., High, A. S., Johnson, C. A., & Robinson, P. A. (2013). Human Homolog of Drosophila Ariadne (HHARI) is a marker of cellular proliferation associated with nuclear bodies. *EXPERIMENTAL CELL RESEARCH*, 319(3), 161-172.

Faucher, D., & Wellinger, R. J. (2010). Methylated H3K4, a transcription-associated histone modification, is involved in the DNA damage response pathway. *PLoS Genetics*, 6(8)

Fei, C., Li, Z., Li, C., Chen, Y., Chen, Z., He, X., Mao, L., Wang, X., Zeng, R., & Li, L. (2013). Smurf1-Mediated Lys29-Linked Nonproteolytic Polyubiquitination of Axin Negatively Regulates Wnt/ -Catenin Signaling. *MOLECULAR AND CELLULAR BIOLOGY*, 33(20), 4095-4105.

Ficarelli, M., Wilson, H., Pedro Galão, R., Mazzon, M., Antzin-Anduetza, I., Marsh, M., Neil, S. J., & Swanson, C. M. (2019). KHNYN is essential for the zinc finger antiviral protein (ZAP) to restrict HIV-1 containing clustered CpG dinucleotides. *eLife*, 8

Flick, K., Raasi, S., Zhang, H., Yen, J. L., & Kaiser, P. (2006). A ubiquitin-interacting motif protects polyubiquitinated Met4 from degradation by the 26S proteasome. *NATURE CELL BIOLOGY*, 8(5), 509-515.

Fraile, J. M., Quesada, V., Rodriguez, D., Freije, J. M., & Lopez-Otin, C. (2012). Deubiquitinases in cancer: new functions and therapeutic options. *ONCOGENE*, 31(19), 2373-2388.

Freedman, D. A., Wu, L., & Levine, A. J. (1999). Functions of the MDM2 oncoprotein. *CELLULAR AND MOLECULAR LIFE SCIENCES*, 55(1), 96-107.

Freyer, M. W., & Lewis, E. A. (2008). Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. *Methods Cell Biol*, 84, 79-113.

Fujita, H., Tokunaga, A., Shimizu, S., Whiting, A. L., Aguilar-Alonso, F., Takagi, K., Walinda, E., Sasaki, Y., Shimokawa, T., Mizushima, T., Ohki, I., Ariyoshi, M., Tochio, H., Bernal, F.,

Shirakawa, M., & Iwai, K. (2018). Cooperative Domain Formation by Homologous Motifs in HOIL-1L and SHARPIN Plays A Crucial Role in LUBAC Stabilization. *Cell Reports*, 23(4), 1192-1204.

Garner, T. P., Strachan, J., Shedden, E. C., Long, J. E., Cavey, J. R., Shaw, B., Layfield, R., & Searle, M. S. (2011). Independent interactions of ubiquitin-binding domains in a ubiquitin-mediated ternary complex. *BIOCHEMISTRY*, 50(42), 9076-9087.

Garnett, M. J., Mansfeld, J., Godwin, C., Matsusaka, T., Wu, J., Russell, P., Pines, J., & Venkitaraman, A. R. (2009). UBE2S elongates ubiquitin chains on APC/C substrates to promote mitotic exit. *NATURE CELL BIOLOGY*, 11(11), 1363-1369.

Gatti, M., Pinato, S., Maiolica, A., Rocchio, F., Prato, M. G., Aebersold, R., & Penengo, L. (2015). RNF168 Promotes Noncanonical K27 Ubiquitination to Signal DNA Damage. *Cell Reports*, 10(2), 226-238.

Geisler, S., Holmström, K. M., Skujat, D., Fiesel, F. C., Rothfuss, O. C., Kahle, P. J., & Springer, W. (2010). PINK1/Parkin-mediated mitophagy is dependent on VDAC1 and p62/SQSTM1. *NATURE CELL BIOLOGY*, 12(2), 119-131.

Gerlach, B., Cordier, S. M., Schmukle, A. C., Emmerich, C. H., Rieser, E., Haas, T. L., Webb, A. I., Rickard, J. A., Anderton, H., Wong, W. W. L., Nachbur, U., Gangoda, L., Warnken, U., Purcell, A. W., Silke, J., & Walczak, H. (2011). Linear ubiquitination prevents inflammation and regulates immune signalling. *NATURE*, 471(7340), 591-596.

Goldstein, G., Scheid, M., Hammerling, U., Schlesinger, D. H., Niall, H. D., & Boyse, E. A. (1975). Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells. *Proc Natl Acad Sci U S A*, 72(1), 11-15.

Gostissa, M., Hengstermann, A., Fogal, V., Sandy, P., Schwarz, S. E., Scheffner, M., & Del, S. G. (1999). Activation of p53 by conjugation to the ubiquitin-like protein SUMO-1. *EMBO JOURNAL*, 18(22), 6462-6471.

Grabbe, C., & Dikic, I. (2009). Functional roles of ubiquitin-like domain (ULD) and ubiquitin-binding domain (UBD) containing proteins. *CHEMICAL REVIEWS*, 109(4), 1481-1494.

Gyrd-Hansen, M., & Meier, P. (2010). IAPs: from caspase inhibitors to modulators of NF-kappaB, inflammation and cancer. *NATURE REVIEWS CANCER*, 10(8), 561-574.

Haas, T. L., Emmerich, C. H., Gerlach, B., Schmukle, A. C., Cordier, S. M., Rieser, E., Feltham, R., Vince, J., Warnken, U., Wenger, T., Koschny, R., Komander, D., Silke, J., & Walczak, H. (2009). Recruitment of the linear ubiquitin chain assembly complex stabilizes the TNF-R1 signaling complex and is required for TNF-mediated gene induction. *MOLECULAR CELL*, 36(5), 831-844.

Hariri nia, A., Verma, R., Purohit, N., Twarog, M. Z., Deshaies, R. J., Bolon, D., & Fushman, D. (2008). Mutations in the hydrophobic core of ubiquitin differentially affect its recognition by receptor proteins. *JOURNAL OF MOLECULAR BIOLOGY*, 375(4), 979-996.

Hirano, S., Kawasaki, M., Ura, H., Kato, R., Raiborg, C., Stenmark, H., & Wakatsuki, S. (2006). Double-sided ubiquitin binding of Hrs-UIM in endosomal protein sorting. *NATURE STRUCTURAL & MOLECULAR BIOLOGY*, 13(3), 272-277.

Hoesel, B., & Schmid, J. A. (2013). The complexity of NF-kappaB signaling in inflammation and cancer. *Molecular Cancer*, 12, 86.

Hofmann, K., & Bucher, P. (1996). The UBA domain: a sequence motif present in multiple enzyme classes of the ubiquitination pathway. *TRENDS IN BIOCHEMICAL SCIENCES*, 21(5), 172-173.

Huang, H., Jeon, M. S., Liao, L., Yang, C., Elly, C., Yates, J. R., & Liu, Y. C. (2010). K33-linked polyubiquitination of T cell receptor-zeta regulates proteolysis-independent T cell signaling. *IMMUNITY*, 33(1), 60-70.

Huang, L., Kinnucan, E., Wang, G., Beaudenon, S., Howley, P. M., Huibregtse, J. M., & Pavletich, N. P. (1999). Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. *SCIENCE*, 286(5443), 1321-1326.

Ikeda, F., Deribe, Y. L., Skanland, S. S., Stieglitz, B., Grabbe, C., Franz-Wachtel, M., van Wijk, S. J., Goswami, P., Nagy, V., Terzic, J., Tokunaga, F., Androulidaki, A., Nakagawa, T., Pasparakis, M., Iwai, K., Sundberg, J. P., Schaefer, L., Rittinger, K., Macek, B., & Dikic, I. (2011). SHARPIN forms a linear ubiquitin ligase complex regulating NF-kappaB activity and apoptosis. *NATURE*, 471(7340), 637-641.

Jeon, Y. J., Yoo, H. M., & Chung, C. H. (2010). ISG15 and immune diseases. *Biochim Biophys Acta*, 1802(5), 485-496.

Jin, L., Williamson, A., Banerjee, S., Philipp, I., & Rape, M. (2008). Mechanism of ubiquitin-chain formation by the human anaphase-promoting complex. *CELL*, 133(4), 653-665.

Kaiser, P., Flick, K., Wittenberg, C., & Reed, S. I. (2000). Regulation of transcription by ubiquitination without proteolysis: Cdc34/SCF(Met30)-mediated inactivation of the transcription factor Met4. *CELL*, 102(3), 303-314.

Kamadurai, H. B., Souphron, J., Scott, D. C., Duda, D. M., Miller, D. J., Stringer, D., Piper, R. C., & Schulman, B. A. (2009). Insights into Ubiquitin Transfer Cascades from a Structure of a UbcH5B~Ubiquitin-HECTNEDD4L Complex. *MOLECULAR CELL*, 36(6), 1095-1102.

Kanayama, A., Seth, R. B., Sun, L., Ea, C. K., Hong, M., Shaito, A., Chiu, Y. H., Deng, L., & Chen, Z. J. (2004). TAB2 and TAB3 activate the NF-kappaB pathway through binding to polyubiquitin chains. *MOLECULAR CELL*, 15(4), 535-548.

Kang, R. S., Daniels, C. M., Francis, S. A., Shih, S. C., Salerno, W. J., Hicke, L., & Radhakrishnan, I. (2003). Solution Structure of a CUE-Ubiquitin Complex Reveals a Conserved Mode of Ubiquitin Binding. *CELL*, 113(5), 621-630.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*, 10(6), 845-858.

Kelsall, I. R., Duda, D. M., Olszewski, J. L., Hofmann, K., Knebel, A., Langevin, F., Wood, N., Wightman, M., Schulman, B. A., & Alpi, A. F. (2013). TRIAD1 and HHARI bind to and are activated by distinct neddylated Cullin-RING ligase complexes. *EMBO JOURNAL*, 32(21), 2848-2860.

Kirisako, T., Kamei, K., Murata, S., Kato, M., Fukumoto, H., Kanie, M., Sano, S., Tokunaga, F., Tanaka, K., & Iwai, K. (2006). A ubiquitin ligase complex assembles linear polyubiquitin chains. *EMBO JOURNAL*, 25(20), 4877-4887.

Kirkpatrick, D. S., Hathaway, N. A., Hanna, J., Elsasser, S., Rush, J., Finley, D., King, R. W., & Gygi, S. P. (2006). Quantitative analysis of in vitro ubiquitinated cyclin B1 reveals complex chain topology. *NATURE CELL BIOLOGY*, 8(7), 700-710.

Kliza, K., & Husnjak, K. (2020). Resolving the Complexity of Ubiquitin Networks. *Front Mol Biosci*, 7, 21.

Komander, D., & Rape, M. (2012). The ubiquitin code. *Annual Review of Biochemistry*, 81, 203-229.

Kulathu, Y., Akutsu, M., Bremm, A., Hofmann, K., & Komander, D. (2009). Two-sided ubiquitin binding explains specificity of the TAB2 NZF domain. *NATURE STRUCTURAL & MOLECULAR BIOLOGY*, 16(12), 1328-1330.

Lawrence, T. (2009). The Nuclear Factor NF- B Pathway in Inflammation. *Cold Spring Harbor Perspectives in Biology*, 1(6), a1651.

Lechtenberg, B. C., Rajput, A., Sanishvili, R., Dobaczewska, M. K., Ware, C. F., Mace, P. D., & Riedl, S. J. (2016). Structure of a HOIP/E2~ubiquitin complex reveals RBR E3 ligase mechanism and regulation. *NATURE*, 529(7587), 546-550.

Lee, S., Tsai, Y. C., Mattera, R., Smith, W. J., Kostelansky, M. S., Weissman, A. M., Bonifacino, J. S., & Hurley, J. H. (2006). Structural basis for ubiquitin recognition and autoubiquitination by Rabex-5. *NATURE STRUCTURAL & MOLECULAR BIOLOGY*, 13(3), 264-271.

Li, M., Brooks, C. L., Wu-Baer, F., Chen, D., Baer, R., & Gu, W. (2003). Mono- versus polyubiquitination: differential control of p53 fate by Mdm2. *SCIENCE*, 302(5652), 1972-1975.

Li, Z., Wang, Y., Li, Y., Yin, W., Mo, L., Qian, X., Zhang, Y., Wang, G., Bu, F., Zhang, Z., Ren, X., Zhu, B., Niu, C., Xiao, W., & Zhang, W. (2018). Ube2s stabilizes beta-Catenin through K11-linked polyubiquitination to promote mesendoderm specification and colorectal cancer development. *Cell Death & Disease*, 9(5), 456.

Liang, J., Saad, Y., Lei, T., Wang, J., Qi, D., Yang, Q., Kolattukudy, P. E., & Fu, M. (2010). MCP-induced protein 1 deubiquitinates TRAF proteins and negatively regulates JNK and NF- κ B signaling. *The Journal of Experimental Medicine*, 207(13), 2959-2973.

Liang, J., Wang, J., Azfer, A., Song, W., Tromp, G., Kolattukudy, P. E., & Fu, M. (2008). A Novel CCCH-Zinc Finger Protein Family Regulates Proinflammatory Activation of Macrophages. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 283(10), 6337-6346.

Liu, J., Wang, Y., Gong, Y., Fu, T., Hu, S., Zhou, Z., & Pan, L. (2017). Structural Insights into SHARPIN-Mediated Activation of HOIP for the Linear Ubiquitin Chain Assembly. *Cell Reports*, 21(1), 27-36.

Liu, S., Chen, Y., Li, J., Huang, T., Tarasov, S., King, A., Weissman, A. M., Byrd, R. A., & Das, R. (2012). Promiscuous interactions of gp78 E3 ligase CUE domain with polyubiquitin chains. *STRUCTURE*, 20(12), 2138-2150.

Liu, S., & Chen, Z. J. (2011). Expanding role of ubiquitination in NF-kappaB signaling. *CELL RESEARCH*, 21(1), 6-21.

Lowe, E. D., Hasan, N., Trempe, J., Fonso, L., Noble, M. E. M., Endicott, J. A., Johnson, L. N., & Brown, N. R. (2006). Structures of the Dsk2 UBL and UBA domains and their complex. *Acta Crystallographica Section D Biological Crystallography*, 62(2), 177-188.

Lutz, J., Höllmüller, E., Scheffner, M., Marx, A., & Stengel, F. (2020). The length of a ubiquitin chain: A general factor for selective recognition by ubiquitin-binding proteins. *Angewandte Chemie*, 132(30), 12470-12474.

Matsushita, K., Takeuchi, O., Standley, D. M., Kumagai, Y., Kawagoe, T., Miyake, T., Satoh, T., Kato, H., Tsujimura, T., Nakamura, H., & Akira, S. (2009). Zc3h12a is an RNase essential for controlling immune responses by regulating mRNA decay. *NATURE*, 458(7242), 1185-1190.

Metzger, M. B., Pruneda, J. N., Klevit, R. E., & Weissman, A. M. (2014). RING-type E3 ligases: Master manipulators of E2 ubiquitin-conjugating enzymes and ubiquitination. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1843(1), 47-60.

Michel, M. A., Swatek, K. N., Hospenthal, M. K., & Komander, D. (2017). Ubiquitin Linkage-Specific Affimers Reveal Insights into K6-Linked Ubiquitin Signaling. *MOLECULAR CELL*, 68(1), 233-246.

Morreale, F. E., & Walden, H. (2016). Types of Ubiquitin Ligases. *CELL*, 165(1), 248.

Morris, J. R., & Solomon, E. (2004). BRCA1 : BARD1 induces the formation of conjugated ubiquitin structures, dependent on K6 of ubiquitin, in cells during DNA replication and repair. *HUMAN MOLECULAR GENETICS*, 13(8), 807-817.

Murillas, R., Simms, K. S., Hatakeyama, S., Weissman, A. M., & Kuehn, M. R. (2002). Identification of Developmentally Expressed Proteins That Functionally Interact with Nedd4 Ubiquitin Ligase. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 277(4), 2897-2907.

Nepravishta, R., Ferrentino, F., Mandaliti, W., Mattioni, A., Weber, J., Polo, S., Castagnoli, L., Cesareni, G., Paci, M., & Santonico, E. (2019). CoCUN, a Novel Ubiquitin Binding Domain Identified in N4BP1. *Biomolecules*, 9(7), 284.

Oberst, A., Malatesta, M., Aqeilan, R. I., Rossi, M., Salomoni, P., Murillas, R., Sharma, P., Kuehn, M. R., Oren, M., Croce, C. M., Bernassola, F., & Melino, G. (2007). The Nedd4-binding partner 1 (N4BP1) protein is an inhibitor of the E3 ligase Itch. *Proc Natl Acad Sci U S A*, 104(27), 11280-11285.

Ohi, M. D., Vander Kooi, C. W., Rosenberg, J. A., Chazin, W. J., & Gould, K. L. (2003). Structural insights into the U-box, a domain associated with multi-ubiquitination. *NATURE STRUCTURAL & MOLECULAR BIOLOGY*, 10(4), 250-255.

Ohtake, F., Tsuchiya, H., Saeki, Y., & Tanaka, K. (2018). K63 ubiquitylation triggers proteasomal degradation by seeding branched ubiquitin chains. *Proceedings of the National Academy of Sciences*, 115(7), E1401-E1408.

Ordureau, A., Heo, J., Duda, D. M., Paulo, J. A., Olszewski, J. L., Yanishevski, D., Rinehart, J., Schulman, B. A., & Harper, J. W. (2015). Defining roles of PARKIN and ubiquitin phosphorylation by PINK1 in mitochondrial quality control using a ubiquitin replacement strategy. *Proceedings of the National Academy of Sciences*, 112(21), 6637-6642.

Pamukcu, B., Lip, G. Y., & Shantsila, E. (2011). The nuclear factor--kappa B pathway in atherosclerosis: a potential therapeutic target for atherothrombotic vascular disease. *THROMBOSIS RESEARCH*, 128(2), 117-123.

Pickart, C. M. (2000). Ubiquitin in chains. *TRENDS IN BIOCHEMICAL SCIENCES*, 25(11), 544-548.

Pickrell, A. M., & Youle, R. J. (2015). The roles of PINK1, parkin, and mitochondrial fidelity in Parkinson's disease. *NEURON*, 85(2), 257-273.

Piñeiro, Á., Muñoz, E., Sabín, J., Costas, M., Bastos, M., Velázquez-Campoy, A., ... & Cotelo, C. (2019). AFFINImeter: A software to analyze molecular recognition processes from experimental data. *Analytical biochemistry*, 577, 117-134.

Plechanovova, A., Jaffray, E. G., Tatham, M. H., Naismith, J. H., & Hay, R. T. (2012). Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. *NATURE*, 489(7414), 115-120.

Ponting, C. P. (2000). Proteins of the endoplasmic-reticulum-associated degradation pathway: domain detection and function prediction. *BIOCHEMICAL JOURNAL*, 351 Pt 2, 527-535.

Popovic, D., Vucic, D., & Dikic, I. (2014). Ubiquitination in disease pathogenesis and treatment. *NATURE MEDICINE*, 20(11), 1242-1253.

Racine, A., Pagé, V., Nagy, S., Grabowski, D., & Tanny, J. C. (2012). Histone H2B Ubiquitylation Promotes Activity of the Intact Set1 Histone Methyltransferase Complex in Fission Yeast. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 287(23), 19040-19047.

Rahighi, S., Ikeda, F., Kawasaki, M., Akutsu, M., Suzuki, N., Kato, R., Kensche, T., Uejima, T., Bloor, S., Komander, D., Randow, F., Wakatsuki, S., & Dikic, I. (2009). Specific recognition of linear ubiquitin chains by NEMO is important for NF-kappaB activation. *CELL*, 136(6), 1098-1109.

Rahighi, S., & Dikic, I. (2012). Selectivity of the ubiquitin-binding modules. *FEBS LETTERS*, 586(17), 2705-2710.

Ramakrishna, S., Suresh, B., & Baek, K. H. (2011). The role of deubiquitinating enzymes in apoptosis. *CELLULAR AND MOLECULAR LIFE SCIENCES*, 68(1), 15-26.

Reid, L. J., Shakya, R., Modi, A. P., Lokshin, M., Cheng, J. T., Jasin, M., Baer, R., & Ludwig, T. (2008). E3 ligase activity of BRCA1 is not essential for mammalian cell viability or homology-directed repair of double-strand DNA breaks. *Proc Natl Acad Sci U S A*, 105(52), 20876-20881.

Reyes-Turcu, F. E., Horton, J. R., Mullally, J. E., Heroux, A., Cheng, X., & Wilkinson, K. D. (2006). The ubiquitin binding domain ZnF UBP recognizes the C-terminal diglycine motif of unanchored ubiquitin. *CELL*, 124(6), 1197-1208.

Reyes-Turcu, F. E., Ventii, K. H., & Wilkinson, K. D. (2009). Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes. *Annual Review of Biochemistry*, 78, 363-397.

Rittinger, K., & Ikeda, F. (2017). Linear ubiquitin chains: enzymes, mechanisms and biology. *Open Biology*, 7(4).

Rohl, C. A., Strauss, C. E., Misura, K. M., & Baker, D. (2004). Protein structure prediction using Rosetta. *Methods in enzymology* (Vol. 383, pp. 66-93). Academic Press.

Santonico, E., Nepravishta, R., Mandaliti, W., Castagnoli, L., Cesareni, G., & Paci, M. (2019). CUBAN, a Case Study of Selective Binding: Structural Details of the Discrimination between Ubiquitin and NEDD8. *INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES*, 20(5), 1185.

Sato, Y., Fujita, H., Yoshikawa, A., Yamashita, M., Yamagata, A., Kaiser, S. E., Iwai, K., & Fukai, S. (2011). Specific recognition of linear ubiquitin chains by the Npl4 zinc finger (NZF) domain of the HOIL-1L subunit of the linear ubiquitin chain assembly complex. *Proceedings of the National Academy of Sciences*, 108(51), 20520-20525.

Sato, Y., Yoshikawa, A., Mimura, H., Yamashita, M., Yamagata, A., & Fukai, S. (2009). Structural basis for specific recognition of Lys 63-linked polyubiquitin chains by tandem UIMs of RAP80. *EMBO JOURNAL*, 28(16), 2461-2468.

Sato, Y., Yoshikawa, A., Yamagata, A., Mimura, H., Yamashita, M., Ookata, K., Nureki, O., Iwai, K., Komada, M., & Fukai, S. (2008). Structural basis for specific cleavage of Lys 63-linked polyubiquitin chains. *NATURE*, 455(7211), 358-362.

Schulman, B. A., & Harper, J. W. (2009). Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. *Nat Rev Mol Cell Biol*, 10(5), 319-331.

Sharma, P., Murillas, R., Zhang, H., & Kuehn, M. R. (2010). N4BP1 is a newly identified nucleolar protein that undergoes SUMO-regulated polyubiquitylation and proteasomal turnover at promyelocytic leukemia nuclear bodies. *JOURNAL OF CELL SCIENCE*, 123(Pt 8), 1227-1234.

Shi, C. S., & Kehrl, J. H. (2010). TRAF6 and A20 regulate lysine 63-linked ubiquitination of Beclin-1 to control TLR4-induced autophagy. *Science Signaling*, 3(123), a42.

Shih, S. C., Prag, G., Francis, S. A., Sutanto, M. A., Hurley, J. H., & Hicke, L. (2003). A ubiquitin-binding motif required for intramolecular monoubiquitylation, the CUE domain. *EMBO JOURNAL*, 22(6), 1273-1281.

Sloper-Mould, K. E., Jemc, J. C., Pickart, C. M., & Hicke, L. (2001). Distinct functional surface regions on ubiquitin. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 276(32), 30483-30489.

Smit, J. J., van Dijk, W. J., El Atmioui, D., Merkx, R., Ovaa, H., & Sixma, T. K. (2013). Target specificity of the E3 ligase LUBAC for ubiquitin and NEMO relies on different minimal requirements. *Journal of Biological Chemistry*, 288(44), 31728-31737.

Song, G., Liu, B., Li, Z., Wu, H., Wang, P., Zhao, K., Jiang, G., Zhang, L., & Gao, C. (2016). E3 ubiquitin ligase RNF128 promotes innate antiviral immunity through K63-linked ubiquitination of TBK1. *NATURE IMMUNOLOGY*, 17(12), 1342-1351.

Spel, L., Nieuwenhuis, J., Haarsma, R., Stickel, E., Bleijerveld, O. B., Altelaar, M., Boelens, J. J., Brummelkamp, T. R., Nierkens, S., & Boes, M. (2018). Nedd4-Binding Protein 1 and TNFAIP3-Interacting Protein 1 Control MHC-1 Display in Neuroblastoma. *CANCER RESEARCH*, 78(23), 6621-6631.

Spit, M., Rieser, E., & Walczak, H. (2019). Linear ubiquitination at a glance. *JOURNAL OF CELL SCIENCE*, 132(2)

Stieglitz, B., Morris-Davies, A. C., Koliopoulos, M. G., Christodoulou, E., & Rittinger, K. (2012). LUBAC synthesizes linear ubiquitin chains via a thioester intermediate. *EMBO REPORTS*, 13(9), 840-846.

Stieglitz, B., Rana, R. R., Koliopoulos, M. G., Morris-Davies, A. C., Schaeffer, V., Christodoulou, E., Howell, S., Brown, N. R., Dikic, I., & Rittinger, K. (2013). Structural basis for ligase-specific conjugation of linear ubiquitin chains by HOIP. *NATURE*, 503(7476), 422-426.

Sun, Z. W., & Allis, C. D. (2002). Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *NATURE*, 418(6893), 104-108.

Thrower, J. S., Hoffman, L., Rechsteiner, M., & Pickart, C. M. (2000). Recognition of the polyubiquitin proteolytic signal. *EMBO JOURNAL*, 19(1), 94-102.

Tokunaga, F., Nakagawa, T., Nakahara, M., Saeki, Y., Taniguchi, M., Sakata, S., Tanaka, K., Nakano, H., & Iwai, K. (2011). SHARPIN is a component of the NF-kappaB-activating linear ubiquitin chain assembly complex. *NATURE*, 471(7340), 633-636.

Trempe, J. F., Brown, N. R., Noble, M. E., & Endicott, J. A. (2010). A new crystal form of Lys48-linked diubiquitin. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, 66(Pt 9), 994-998.

Tsuchiya, H., Burana, D., Ohtake, F., Arai, N., Kaiho, A., Komada, M., ... & Saeki, Y. (2018). Ub-ProT reveals global length and composition of protein ubiquitylation in cells. *Nature communications*, 9(1), 1-10.

Uehata, T., Iwasaki, H., Vandenbon, A., Matsushita, K., Hernandez-Cuellar, E., Kuniyoshi, K., Satoh, T., Mino, T., Suzuki, Y., Standley, D. M., Tsujimura, T., Rakugi, H., Isaka, Y., Takeuchi, O., & Akira, S. (2013). Malt1-Induced Cleavage of Regnase-1 in CD4+ Helper T Cells Regulates Immune Activation. *CELL*, 153(5), 1036-1049.

van der Veen, A. G., & Ploegh, H. L. (2012). Ubiquitin-like proteins. *Annual Review of Biochemistry*, 81, 323-357.

Varadan, R., Assfalg, M., Raasi, S., Pickart, C., & Fushman, D. (2005). Structural determinants for selective recognition of a Lys48-linked polyubiquitin chain by a UBA domain. *MOLECULAR CELL*, 18(6), 687-698.

Vives-Bauza, C., Zhou, C., Huang, Y., Cui, M., de Vries, R. L., Kim, J., May, J., Tocilescu, M. A., Liu, W., Ko, H. S., Magrane, J., Moore, D. J., Dawson, V. L., Grailhe, R., Dawson, T. M., Li, C., Tieu, K., & Przedborski, S. (2010). PINK1-dependent recruitment of Parkin to mitochondria in mitophagy. *Proc Natl Acad Sci U S A*, 107(1), 378-383.

von Stechow, L., Typas, D., Carreras, P. J., Oort, L., Siddappa, R., Pines, A., Vrieling, H., van de Water, B., Mullenders, L. H., & Danen, E. H. (2015). The E3 ubiquitin ligase ARIH1 protects against genotoxic stress by initiating a 4EHP-mediated mRNA translation arrest. *MOLECULAR AND CELLULAR BIOLOGY*, 35(7), 1254-1268.

Walters, K. J., Goh, A. M., Wang, Q., Wagner, G., & Howley, P. M. (2004). Ubiquitin family proteins and their relationship to the proteasome: a structural perspective. *Biochim Biophys Acta*, 1695(1-3), 73-87.

Wang, B., Alam, S. L., Meyer, H. H., Payne, M., Stemmler, T. L., Davis, D. R., & Sundquist, W. I. (2003). Structure and Ubiquitin Interactions of the Conserved Zinc Finger Domain of Npl4. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 278(22), 20225-20234.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S., & Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *NATURE*, 431(7010), 873-878.

Wang, Q., Huang, L., Hong, Z., Lv, Z., Mao, Z., Tang, Y., Kong, X., Li, S., Cui, Y., Liu, H., Zhang, L., Zhang, X., Jiang, L., Wang, C., & Zhou, Q. (2017). The E3 ubiquitin ligase RNF185 facilitates the cGAS-mediated innate immune response. *PLoS Pathogens*, 13(3), e1006264.

Wang, Q., Liu, X., Cui, Y., Tang, Y., Chen, W., Li, S., Yu, H., Pan, Y., & Wang, C. (2014). The E3 Ubiquitin Ligase AMFR and INSIG1 Bridge the Activation of TBK1 Kinase by Modifying the Adaptor STING. *IMMUNITY*, 41(6), 919-933.

Waudby, C. A., Ramos, A., Cabrita, L. D., & Christodoulou, J. (2016). Two-Dimensional NMR Lineshape Analysis. *Scientific Reports*, 6(1)

Wauer, T., & Komander, D. (2013). Structure of the human Parkin ligase domain in an autoinhibited state. *The EMBO journal*, 32(15), 2099-2112.

Weake, V. M., & Workman, J. L. (2008). Histone ubiquitination: triggering gene activity. *MOLECULAR CELL*, 29(6), 653-663.

Weeks, S. D., Grasty, K. C., Hernandez-Cuebas, L., & Loll, P. J. (2009). Crystal structures of Lys-63-linked tri- and di-ubiquitin reveal a highly extended chain architecture. *PROTEINS-STRUCTURE FUNCTION AND BIOINFORMATICS*, 77(4), 753-759.

Williamson, M. P. (2013). Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc*, 73, 1-16.

Wu-Baer, F., Lagrazon, K., Yuan, W., & Baer, R. (2003). The BRCA1/BARD1 Heterodimer Assembles Polyubiquitin Chains through an Unconventional Linkage Involving Lysine Residue K6 of Ubiquitin. *The Journal of biological chemistry*, 278(37), 34743-34746.

Xirodimas, D. P., Saville, M. K., Bourdon, J. C., Hay, R. T., & Lane, D. P. (2004). Mdm2-mediated NEDD8 conjugation of p53 inhibits its transcriptional activity. *CELL*, 118(1), 83-97.

Xu, J., Peng, W., Sun, Y., Wang, X., Xu, Y., Li, X., Gao, G., & Rao, Z. (2012). Structural study of MCPIP1 N-terminal conserved domain reveals a PIN-like RNase. *NUCLEIC ACIDS RESEARCH*, 40(14), 6957-6965.

Yagi, H., Ishimoto, K., Hiromoto, T., Fujita, H., Mizushima, T., Uekusa, Y., Yagi-Utsumi, M., Kurimoto, E., Noda, M., Uchiyama, S., Tokunaga, F., Iwai, K., & Kato, K. (2012). A

non-canonical UBA-UBL interaction forms the linear-ubiquitin-chain assembly complex. *EMBO REPORTS*, 13(5), 462-468.

Yamasoba, D., Sato, K., Ichinose, T., Imamura, T., Koepke, L., Joas, S., Reith, E., Hotter, D., Misawa, N., Akaki, K., Uehata, T., Mino, T., Miyamoto, S., Noda, T., Yamashita, A., Standley, D. M., Kirchhoff, F., Sauter, D., Koyanagi, Y., & Takeuchi, O. (2019). N4BP1 restricts HIV-1 and its inactivation by MALT1 promotes viral reactivation. *Nature Microbiology*, 4(9), 1532-1544.

Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., & Baker, D. (2020). Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences*, 117(3), 1496-1503.

Yang, M., Chen, T., Li, X., Yu, Z., Tang, S., Wang, C., Gu, Y., Liu, Y., Xu, S., Li, W., Zhang, X., Wang, J., & Cao, X. (2015). K33-linked polyubiquitination of Zap70 by Nrdp1 controls CD8(+) T cell activation. *NATURE IMMUNOLOGY*, 16(12), 1253-1262.

Yang, W. L., Wang, J., Chan, C. H., Lee, S. W., Campos, A. D., Lamothe, B., Hur, L., Grabiner, B. C., Lin, X., Darnay, B. G., & Lin, H. K. (2009). The E3 Ligase TRAF6 Regulates Akt Ubiquitination and Activation. *SCIENCE*, 325(5944), 1134-1138.

Ye, Y., & Rape, M. (2009). Building ubiquitin chains: E2 enzymes at work. *Nat Rev Mol Cell Biol*, 10(11), 755-764.

Yeh, E. T., Gong, L., & Kamitani, T. (2000). Ubiquitin-like proteins: new wines in new bottles. *GENE*, 248(1-2), 1-14.

Yokogawa, M., Tsushima, T., Noda, N. N., Kumeta, H., Enokizono, Y., Yamashita, K., Standley, D. M., Takeuchi, O., Akira, S., & Inagaki, F. (2016). Structural basis for the regulation of enzymatic activity of Regnase-1 by domain-domain interactions. *Sci Rep*, 6, 22324.

Yuan, L., Lv, Z., Atkison, J. H., & Olsen, S. K. (2017). Structural insights into the mechanism and E2 specificity of the RBR E3 ubiquitin ligase HHARI. *Nature Communications*, 8(1), 211.

Yuan, W. C., Lee, Y. R., Lin, S. Y., Chang, L. Y., Tan, Y. P., Hung, C. C., Kuo, J. C., Liu, C. H., Lin, M. Y., Xu, M., Chen, Z. J., & Chen, R. H. (2014). K33-Linked Polyubiquitination of Coronin 7 by Cul3-KLHL20 Ubiquitin E3 Ligase Regulates Protein Trafficking. *MOLECULAR CELL*, 54(4), 586-600.

Zheng, N., Schulman, B. A., Song, L., Miller, J. J., Jeffrey, P. D., Wang, P., Chu, C., Koepp, D. M., Elledge, S. J., Pagano, M., Conaway, R. C., Conaway, J. W., Harper, J. W., & Pavletich, N. P. (2002). Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. *NATURE*, 416(6882), 703-709.

Zheng, N., & Shabek, N. (2017). Ubiquitin Ligases: Structure, Function, and Regulation. *Annual Review of Biochemistry*, 86, 129-157.

Zhou, L., Azfer, A., Niu, J., Graham, S., Choudhury, M., Adamski, F. M., Younce, C., Binkley, P. F., & Kolattukudy, P. E. (2006). Monocyte chemoattractant protein-1 induces a novel transcription factor that causes cardiac myocyte apoptosis and ventricular dysfunction. *CIRCULATION RESEARCH*, 98(9), 1177-1185.

Zhu, P., Zhou, W., Wang, J., Puc, J., Ohgi, K. A., Erdjument-Bromage, H., Tempst, P., Glass, C. K., & Rosenfeld, M. G. (2007). A histone H2A deubiquitinase complex coordinating histone acetylation and H1 dissociation in transcriptional regulation. *MOLECULAR CELL*, 27(4), 609-621.

Zotti, T., Uva, A., Ferravante, A., Vessichelli, M., Scudiero, I., Ceccarelli, M., Vito, P., & Stilo, R. (2011). TRAF7 Protein Promotes Lys-29-linked Polyubiquitination of I κ B Kinase (IKK γ)/NF-

κ B Essential Modulator (NEMO) and p65/RelA Protein and Represses NF- κ B Activation. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 286(26), 22924-22933.

Appendix I

Full chemical shift assignment list of the N4BP1 CUE domain

Residue		Atom	Chemical	SD
No.	Name	Name	Shift	
850	Glu	H	8.09629	0.00549
850	Glu	N	120.31422	0.06579
850	Glu	Cβ	30.23615	0.02078
850	Glu	Cα	57.04142	0.10474
850	Glu	C	177.39061	0.02378
850	Glu	Hα	4.44084	0.00267
850	Glu	Hβa	2.11863	0.00113
850	Glu	Hβb	2.18288	0.00278
850	Glu	Cγ	36.82023	0
850	Glu	Hγa	2.33323	0.00146
850	Glu	Hγb	2.33323	0.00146
851	Thr	H	8.26442	0.00583
851	Thr	N	115.12059	0.02795
851	Thr	C	174.99923	0.02699
851	Thr	Cα	63.3913	0.0411
851	Thr	Cβ	69.32607	0.07823
851	Thr	Hα	4.31951	5.53E-04
851	Thr	Hβ	4.29038	4.67E-04
852	Ser	N	117.20887	0.04914
852	Ser	H	8.33714	0.00703
852	Ser	Cα	59.93276	0.04293
852	Ser	Cβ	63.21361	0.08294
852	Ser	C	175.59769	0
852	Ser	Hα	4.42433	0.00515
852	Ser	Hβb	4.05257	0.01025
852	Ser	Hβa	3.97813	0.00438
853	Glu	H	8.37014	0.00542
853	Glu	N	123.128	0.07468
853	Glu	Cα	58.83371	0.07559
853	Glu	Cβ	29.83386	0.06249
853	Glu	C	178.74942	0.01861
853	Glu	Hα	4.21881	0.00378
853	Glu	Hβa	2.13675	0.00277
853	Glu	Hβb	2.13687	0.00292
853	Glu	Hγb	2.40338	0.0015

Residue		Atom	Chemical	SD
No.	Name	Name	Shift	
853	Glu	Hγa	2.33514	0.00515
853	Glu	Cγ	36.64704	0
854	Leu	H	7.98784	0.00392
854	Leu	N	121.61625	0.0461
854	Leu	Cα	57.4809	0.03315
854	Leu	Cβ	42.21615	0.02716
854	Leu	C	177.94657	0.06515
854	Leu	Hα	4.18322	0.00376
854	Leu	Hβb	1.80022	0.00497
854	Leu	Hβa	1.51728	0.01133
854	Leu	Cδb	24.89424	3.37E-07
854	Leu	Cδa	24.15242	0
854	Leu	Cγ	26.85965	4.77E-07
854	Leu	Hγ	1.68187	0.00727
854	Leu	Hδa*	0.87789	0.00223
854	Leu	Hδb*	0.92233	0.00277
855	Arg	H	8.12983	0.00553
855	Arg	N	119.05411	0.05761
855	Arg	Cβ	30.14622	0.00891
855	Arg	Cα	60.23006	0.10028
855	Arg	C	177.61017	0.01239
855	Arg	Hα	3.80196	0.00313
855	Arg	Hβb	1.95359	0.00227
855	Arg	Hβa	1.81888	0.00183
855	Arg	Cγ	27.95641	3.37E-07
855	Arg	Hγb	1.59034	0.00228
855	Arg	Hγa	1.48044	0.00175
855	Arg	Cδ	43.54572	4.77E-07
855	Arg	Hδb	3.29597	0.00135
855	Arg	Hδa	3.16435	0.00234
856	Glu	H	7.97112	0.00464
856	Glu	N	116.21697	0.06065
856	Glu	Cα	59.14594	0.05624
856	Glu	Cβ	29.50653	0.09037
856	Glu	C	178.79195	0.0246

Residue		Atom	Chemical	SD
No.	Name	Name	Shift	
856	Glu	Hα	3.99803	0.00169
856	Glu	Hβb	2.10851	0.00185
856	Glu	Hβa	2.1084	0.00177
856	Glu	Cγ	36.2916	4.77E-07
856	Glu	Hγb	2.41173	0.00155
856	Glu	Hγa	2.38281	8.50E-04
857	Ala	H	7.80064	0.00184
857	Ala	N	121.094	0.05281
857	Ala	Cα	54.9543	0.02988
857	Ala	Cβ	18.7895	0.05326
857	Ala	C	180.704	0.03011
857	Ala	Hα	4.16664	0.00158
857	Ala	Hβ*	1.57795	0.0027
858	Leu	H	8.20107	0.00522
858	Leu	N	118.795	0.01867
858	Leu	Cα	57.7554	0.05957
858	Leu	Cβ	42.6221	0.01129
858	Leu	C	178.644	0.01153
858	Leu	Hα	4.18868	0.0045
858	Leu	Hβb	2.0874	0.00503
858	Leu	Hβa	1.53785	0.00242
858	Leu	Cγ	26.8942	0
858	Leu	Hγ	1.90272	0.00766
858	Leu	Cδb	24.0427	0
858	Leu	Hδb*	0.98716	0.00208
858	Leu	Cδa	26.6646	0
858	Leu	Hδa*	0.92698	0.00306
859	Leu	H	8.26439	0.00851
859	Leu	N	119.22	0.05008
859	Leu	Cβ	41.8063	0.03857
859	Leu	Cα	56.7925	0.1639
859	Leu	C	178.638	0.00234
859	Leu	Hα	4.15521	0.00241
859	Leu	Hβb	1.88375	0.0071
859	Leu	Hβa	1.52466	0.00641
859	Leu	Cδb	23.1468	0
859	Leu	Hδb*	0.94597	9.55E-04
859	Leu	Cδa	25.6006	0
859	Leu	Hδa*	0.91052	0.00201
860	Lys	H	7.29934	0.00352
860	Lys	N	116.238	0.05614
860	Lys	Cα	57.9332	0.08189
860	Lys	Cβ	32.7757	6.74E-07
860	Lys	C	177.84	0.00576
860	Lys	Hα	4.12328	0.00356
860	Lys	Hβb	1.91612	0.00388
860	Lys	Hβa	1.91539	0.00356
860	Lys	Cγ	25.2764	3.37E-07
860	Lys	Hγb	1.64656	0.00149
860	Lys	Hγa	1.47579	0.00481
860	Lys	Cδ	28.8931	4.77E-07
860	Lys	Hδa	1.72046	0.00353
860	Lys	Hδb	1.72046	0.00353
860	Lys	Cε	42.2128	0.00824
860	Lys	Hεa	3.03621	0.00285
860	Lys	Hεb	3.03621	0.00285
861	Ile	H	7.20144	0.00294
861	Ile	N	117.623	0.02096
861	Ile	Cα	62.9815	0.05552
861	Ile	Cβ	39.0492	0.02049
861	Ile	C	175.432	0.00203
861	Ile	Hα	3.8023	0.00587
861	Ile	Hβ	1.59729	0.00222
861	Ile	Hγ1b	1.49252	0.00208
861	Ile	Cγ1	27.5811	0.00346
861	Ile	Hγ1a	1.00397	0.00327
861	Ile	Cγ2	17.2953	0.01505
861	Ile	Hγ2*	0.27291	0.00235
861	Ile	Cδ1	13.5253	0.02146
861	Ile	Hδ1*	0.76578	0.00239
862	Phe	H	7.86912	0.00712
862	Phe	N	119.783	0.0612
862	Phe	Cα	55.2044	0.01902
862	Phe	Cβ	39.1369	0
862	Phe	C	172.737	0
862	Phe	Hα	5.05266	0.00132
862	Phe	Hβb	3.01858	0
862	Phe	Hβa	2.98603	0.00418
863	Pro	Hα	4.68758	0.00428
863	Pro	Hβb	2.27898	0.00357
863	Pro	Hβa	2.08149	0.00395
863	Pro	C	177.901	0
863	Pro	Cα	63.676	0.10253
863	Pro	Cβ	32.9096	4.77E-07

Residue		Atom	Chemical	SD
No.	Name	Name	Shift	
863	Pro	Cγ	26.3509	3.37E-07
863	Pro	Hγa	1.98092	0.00175
863	Pro	Hγb	1.98097	0.00158
863	Pro	Cδ	50.1881	0.01452
863	Pro	Hδb	3.75909	0.0025
863	Pro	Hδa	3.49974	0.00103
864	Asp	H	8.16817	0.00364
864	Asp	N	120.531	0.06676
864	Asp	Cβ	42.1296	0.06683
864	Asp	Cα	54.7433	0.03448
864	Asp	C	177.907	0
864	Asp	Hα	4.73769	0.00132
864	Asp	Hβb	2.87327	0.00346
864	Asp	Hβa	2.78036	0.00461
865	Ser	Hα	4.2059	0.00501
865	Ser	Hβa	4.02295	0.00472
865	Ser	Hβb	4.02295	0.00472
865	Ser	C	176.299	0
865	Ser	Cα	61.8636	0
865	Ser	Cβ	63.4852	0
866	Glu	H	8.86338	0.01399
866	Glu	N	121.114	0.0386
866	Glu	Cβ	29.2864	0.18082
866	Glu	C	178.499	0.01258
866	Glu	Cα	58.9041	0.05924
866	Glu	Hα	4.20821	0
866	Glu	Hβa	2.11296	0
866	Glu	Hβb	2.11296	0
867	Gln	H	7.76158	0.00409
867	Gln	N	117.75	0.05862
867	Gln	Cα	57.4756	0.15808
867	Gln	Cβ	28.4956	0.06969
867	Gln	C	176.949	0.02566
867	Gln	Hα	4.21834	0.00147
867	Gln	Hβb	2.24256	0.0032
867	Gln	Hβa	1.94189	0.0031
867	Gln	Cγ	34.4686	4.77E-07
867	Gln	Hγb	2.41192	0.00198
867	Gln	Hγa	2.33079	0.00239
868	Lys	H	7.7575	0.00587
868	Lys	N	120.152	0.02388
868	Lys	Cα	59.9546	0.09999
868	Lys	Cβ	32.4403	0.0472
868	Lys	C	177.807	0.042
868	Lys	Hα	3.94321	0.00247
868	Lys	Hβa	1.91229	0.00222
868	Lys	Hβb	1.9535	0.0025
868	Lys	Cγ	25.2764	3.37E-07
868	Lys	Hγa	1.46519	0.00542
868	Lys	Hγb	1.55405	0.0064
868	Lys	Cδ	29.6743	5.84E-07
868	Lys	Hδb	1.73357	0.00339
868	Lys	Hδa	1.73315	0.00319
868	Lys	Cε	42.2244	4.77E-07
868	Lys	Hεa	3.03437	0.00297
868	Lys	Hεb	3.03477	0.00282
869	Leu	H	7.5776	0.00398
869	Leu	N	117.685	0.0783
869	Leu	Cα	57.4517	0.15305
869	Leu	Cβ	41.7007	0.03246
869	Leu	C	179.013	0.00721
869	Leu	Hα	4.23806	0.00195
869	Leu	Hβb	1.78746	0.00239
869	Leu	Hβa	1.6478	0.0048
869	Leu	Cγ	27.0504	0
869	Leu	Hγ	1.73922	0.0025
869	Leu	Cδb	24.3551	0
869	Leu	Hδb*	0.98976	0.00126
869	Leu	Cδa	24.0086	0
869	Leu	Hδa*	0.93839	0.00129
870	Lys	H	7.43241	0.00492
870	Lys	N	119.745	0.05829
870	Lys	Cα	58.1664	0.13547
870	Lys	Cβ	32.5915	0.0583
870	Lys	C	178.294	0.00871
870	Lys	Hα	4.18154	0.00429
870	Lys	Hβa	1.93246	0.0047
870	Lys	Hβb	1.93246	0.0047
870	Lys	Cγ	24.964	0
870	Lys	Hγa	1.53745	0.00147
870	Lys	Hγb	1.53765	0.00147
870	Lys	Cδ	29.0493	4.77E-07
870	Lys	Hδa	1.78621	0.0048
870	Lys	Hδb	1.78621	0.0048

Residue		Atom	Chemical	SD
No.	Name	Name	Shift	
870	Lys	Hεa	3.03879	7.64E-04
870	Lys	Hεb	3.03879	7.64E-04
871	Ile	H	8.15904	0.00582
871	Ile	N	119.766	0.05856
871	Ile	Cβ	38.0721	0.05619
871	Ile	Cα	65.5724	0.14795
871	Ile	C	176.878	0.00696
871	Ile	Hα	3.52254	0.00366
871	Ile	Hβ	2.02602	0.00364
871	Ile	Cγ2	17.487	0.03299
871	Ile	Hγ2*	0.92966	0.00353
871	Ile	Cδ1	14.1374	0.02193
871	Ile	Hδ1*	0.92632	0.00563
871	Ile	Cγ1	30.542	0.01963
871	Ile	Hγ1a	0.97757	0.00441
871	Ile	Hγ1b	1.70294	0.0013
872	Asp	H	7.87591	0.0083
872	Asp	N	118.717	0.06038
872	Asp	Cα	57.6551	0.0633
872	Asp	Cβ	40.1209	0.01395
872	Asp	C	179.439	0.02414
872	Asp	Hα	4.49454	0.00222
872	Asp	Hβb	2.75854	4.10E-04
872	Asp	Hβa	2.67651	0.00599
873	Gln	H	7.90065	0.00764
873	Gln	N	119.765	0.05288
873	Gln	Cα	58.7929	0.07978
873	Gln	Cβ	28.6422	0.08718
873	Gln	C	178.979	0.0039
873	Gln	Hα	4.13587	0.00412
873	Gln	Hβb	2.32871	0.00452
873	Gln	Hβa	2.18527	0.00211
873	Gln	Cγ	33.8436	0
873	Gln	Hγb	2.59886	0.00108
873	Gln	Hγa	2.4155	0.00129
874	Ile	H	8.41827	0.01871
874	Ile	N	121.613	0.03438
874	Ile	Cα	64.3688	0.03043
874	Ile	Cβ	37.4898	0.00936
874	Ile	C	178.391	0.03093
874	Ile	Hα	3.95593	0.00216
874	Ile	Hβ	2.03401	0.00355

Residue		Atom	Chemical	SD
No.	Name	Name	Shift	
874	Ile	Cγ1	28.6746	4.77E-07
874	Ile	Hγ1b	1.63895	0.00101
874	Ile	Hγ1a	1.35955	7.10E-04
874	Ile	Cγ2	18.2461	0
874	Ile	Hγ2*	0.92857	8.97E-04
874	Ile	Cδ1	13.8474	0
874	Ile	Hδ1*	0.79977	8.25E-04
875	Leu	H	8.36051	0.00678
875	Leu	N	119.913	0.03797
875	Leu	Cβ	41.0843	0.07007
875	Leu	Cα	57.541	0.05608
875	Leu	C	179.28	0.04123
875	Leu	Hα	4.06523	0.00506
875	Leu	Hβb	1.86484	0.0068
875	Leu	Hβa	1.47942	0.00543
875	Leu	Cδb	22.949	0
875	Leu	Hδb*	0.87416	0.00241
875	Leu	Cδa	25.8834	0
875	Leu	Hδa*	0.87396	0.00196
875	Leu	Cγ	26.5817	0
875	Leu	Hγ	1.87544	0.00334
876	Ala	H	7.51682	0.00778
876	Ala	N	119.165	0.0597
876	Ala	Cα	53.9187	0.08417
876	Ala	Cβ	18.2907	0.16455
876	Ala	C	178.916	0.03093
876	Ala	Hα	4.15281	0.00218
876	Ala	Hβ*	1.47892	0.00192
877	Ala	H	7.54681	0.00774
877	Ala	N	118.421	0.07757
877	Ala	Cα	52.8415	0.06334
877	Ala	Cβ	19.2129	0.01321
877	Ala	C	177.305	0.01031
877	Ala	Hα	4.22297	0.00479
877	Ala	Hβ*	1.31484	0.00119
878	His	H	7.62823	0.00282
878	His	N	115.8	0.1143
878	His	Cα	53.7042	0.07041
878	His	Cβ	29.6642	0.04828
878	His	C	173.164	0
878	His	Hα	4.92971	0.00121
878	His	Hβb	3.0162	0.0017

Residue		Atom	Chemical	SD
No.	Name	Name	Shift	
878	His	Hβa	2.89972	0.0013
879	Pro	C	176.573	0
879	Pro	Cβ	31.4713	5.84E-07
879	Pro	Cα	64.5846	0.06347
879	Pro	Hα	4.44034	0.00192
879	Pro	Hβb	2.22887	0.00332
879	Pro	Hβa	1.70084	0.00223
879	Pro	Cγ	27.2529	6.74E-07
879	Pro	Hγb	1.94806	0.00237
879	Pro	Hγa	1.69754	0.00291
879	Pro	Cδ	50.0597	6.74E-07
879	Pro	Hδb	3.45294	0.01065
879	Pro	Hδa	3.43333	0.00733
880	Tyr	N	115.455	0.10386
880	Tyr	C	175.727	0.03301
880	Tyr	H	7.72745	0.01725
880	Tyr	Cβ	37.8631	0.05164
880	Tyr	Cα	56.3868	0.03244
880	Tyr	Hα	4.69267	0.00526
880	Tyr	Hβb	3.25107	0.00353
880	Tyr	Hβa	3.08055	0.00518
881	Met	H	7.3873	0.01575
881	Met	N	122.519	0.01209
881	Met	C	175.16	0
881	Met	Cβ	33.7922	0.00104
881	Met	Cα	56.1775	0
881	Met	Hα	4.3915	0.00193
881	Met	Hβb	2.10066	0.00143
881	Met	Hβa	2.00082	0.00175
881	Met	Cγ	32.3857	0
881	Met	Hγb	2.49945	0.00147
881	Met	Hγa	2.37641	0.00313
882	Lys	Hα	4.47614	0.00318
882	Lys	Hβb	1.99422	0.00254
882	Lys	Hβa	1.76864	0.00278
882	Lys	C	175.105	0
882	Lys	Cβ	33.2495	0.0184
882	Lys	Cα	55.3267	0.06015
882	Lys	Cγ	24.0265	5.84E-07
882	Lys	Hγb	1.47847	0.00179
882	Lys	Hγa	1.41202	0.00184
882	Lys	Cδ	28.5806	3.37E-07

Residue		Atom	Chemical	SD
No.	Name	Name	Shift	
882	Lys	Hδa	1.71363	0.00219
882	Lys	Hδb	1.71363	0.00219
882	Lys	Cε	42.8494	0
882	Lys	Hεa	3.05549	0.00218
882	Lys	Hεb	3.05569	0.00224
883	Asp	H	7.17861	0.01917
883	Asp	N	118.628	0.10886
883	Asp	Cα	53.4716	0.11584
883	Asp	Cβ	42.274	0.06952
883	Asp	Hα	4.71019	8.11E-04
883	Asp	Hβb	2.87945	0.00528
883	Asp	Hβa	2.70758	0.00543
883	Asp	C	176.091	0.00416
884	Leu	H	8.67368	0.01377
884	Leu	N	126.55	0.08685
884	Leu	Hα	4.06978	0.00669
884	Leu	Hβa	1.6475	0.00455
884	Leu	Hβb	1.71405	0.00167
884	Leu	C	178.926	0.01058
884	Leu	Cα	57.5991	6.74E-07
884	Leu	Cβ	42.1532	6.74E-07
884	Leu	Cγ	26.8942	4.77E-07
884	Leu	Hγ	1.71305	0.00267
884	Leu	Cδa	24.1989	0
884	Leu	Hδa*	0.91592	0.00332
884	Leu	Cδb	25.1022	0
884	Leu	Hδb*	0.95163	0.00179
885	Asn	H	8.43173	0.00931
885	Asn	N	118.696	0.04015
885	Asn	Cα	56.1423	0.11868
885	Asn	Cβ	38.0127	0.01407
885	Asn	C	177.243	0
885	Asn	Hα	4.50167	0.00722
885	Asn	Hβb	2.99602	8.50E-04
885	Asn	Hβa	2.84701	0.00267
885	Asn	Nd2	113.666	0.02465
885	Asn	Hδ2b	7.79014	0.00525
885	Asn	Hδ2a	6.91251	0.00272
886	Ala	H	7.86152	0.002
886	Ala	N	124.32	0.04563
886	Ala	Cα	54.3886	0.05908
886	Ala	Cβ	18.7821	2.38E-07

Residue		Atom	Chemical	SD
No.	Name	Name	Shift	
886	Ala	C	179.992	1.91E-06
886	Ala	Hα	4.21537	0.00604
886	Ala	Hβ*	1.51465	0.00185
887	Leu	H	8.05634	0.00893
887	Leu	N	118.261	0.05979
887	Leu	Cα	57.4633	0.09047
887	Leu	Cβ	43.216	0.05622
887	Leu	C	178.648	0.01234
887	Leu	Hα	4.0626	0.01148
887	Leu	Hβb	1.8814	0.00517
887	Leu	Hβa	1.43823	0.00803
887	Leu	Cγ	27.8491	0
887	Leu	Hγ	1.64082	0.00144
887	Leu	Cδa	23.7302	0
887	Leu	Hδb*	0.86229	0.00165
887	Leu	Cδb	26.3521	0
887	Leu	Hδa*	0.86001	8.50E-04
888	Ser	H	8.23167	0.01184
888	Ser	N	113.876	0.07414
888	Ser	Cβ	63.2108	0.07844
888	Ser	Cα	61.9158	0.11383
888	Ser	Hα	4.19304	0
888	Ser	Hβa	4.00923	0
888	Ser	Hβb	4.00923	0
888	Ser	C	175.32	0
889	Ala	H	7.45268	0.00845
889	Ala	N	122.257	0.07279
889	Ala	Cα	54.2191	0.11425
889	Ala	Cβ	18.3724	0.00342
889	Ala	C	179.035	0.01231
889	Ala	Hα	4.22546	0.00249
889	Ala	Hβ*	1.4988	0.00516
890	Leu	H	7.38504	0.00737
890	Leu	N	117.066	0.10314
890	Leu	Cα	56.1446	0.11266
890	Leu	Cβ	42.3105	0.03678
890	Leu	C	178.295	0
890	Leu	Hα	4.24396	0.00552
890	Leu	Hβb	1.95295	0.00181
890	Leu	Hβa	1.45344	0.00284
890	Leu	Cγ	26.2868	3.37E-07
890	Leu	Hγ	1.75265	0.00421
890	Leu	Cδa	25.7272	3.37E-07
890	Leu	Hδa*	0.62804	0.00212
890	Leu	Cδb	22.7928	3.37E-07
890	Leu	Hδb*	0.82602	0.00179
891	Val	H	7.58846	0.00583
891	Val	N	119.438	0.04858
891	Val	Hα	3.56229	0.02212
891	Val	Hβ	1.8538	0.00776
891	Val	C	176.468	0
891	Val	Cβ	31.7106	0.07969
891	Val	Cα	64.866	0.0286
891	Val	Cγa	22.0939	2.38E-07
891	Val	Cγb	22.0939	2.38E-07
891	Val	Hγa*	0.81864	0.00297
891	Val	Hγb*	0.81918	0.00297
892	Leu	H	7.51931	0.01412
892	Leu	N	121.079	0.15188
892	Leu	Cα	55.157	0.12877
892	Leu	Hα	4.26399	0.0058
892	Leu	Hβa	1.67129	0.00382
892	Leu	C	176.219	0.01051
892	Leu	Cγ	26.7555	4.77E-07
892	Leu	Hγ	1.73772	5.24E-04
892	Leu	Cδb	25.2925	0
892	Leu	Hδb*	1.01228	7.00E-04
892	Leu	Cδa	23.2274	0
892	Leu	Hδa*	0.95517	8.86E-04
892	Leu	Cβ	42.9379	0.21664
892	Leu	Hβb	1.73791	7.10E-04
893	Asp	H	7.47737	0.00388
893	Asp	N	125.456	0.06679
893	Asp	Cα	56.273	0.0295
893	Asp	Cβ	42.4644	4.77E-07
893	Asp	Hα	4.4153	0.0011
893	Asp	Hβb	2.68709	0.00118
893	Asp	Hβa	2.59496	0.00474

* indicates the atoms that are equivalent to each other and thus have the same chemical shift value

Appendix II

Full $^{1}H/^{15}N$-HSQC assignments of diUb chain

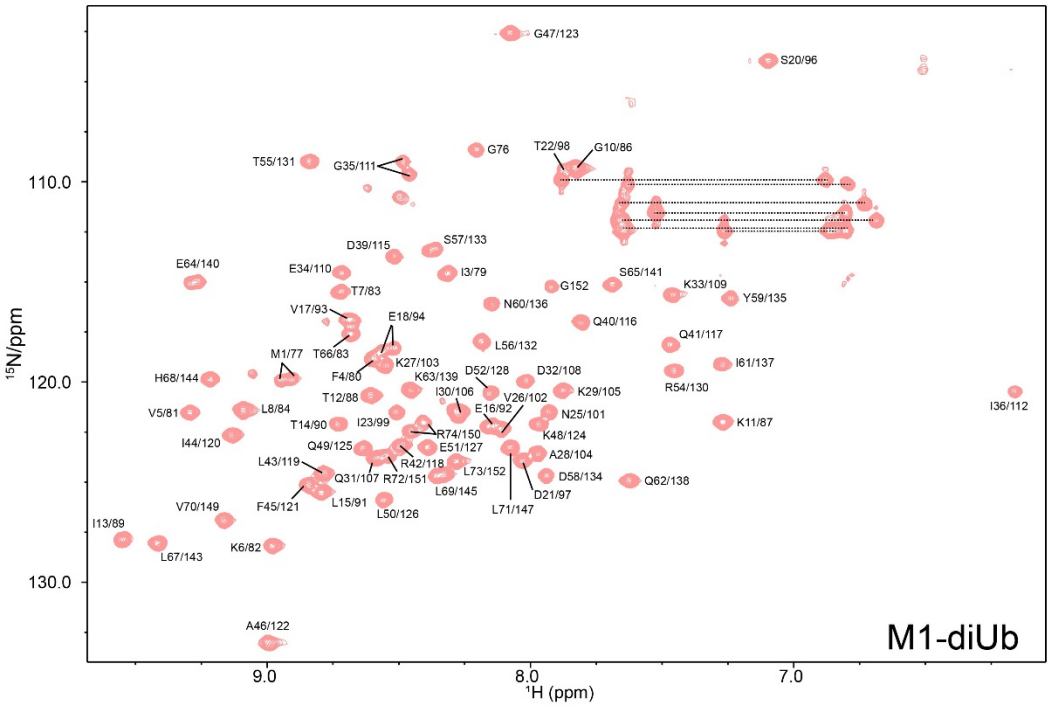


Figure A1 Full $^{1}H/^{15}N$-HSQC assignments of M1-diUb. The spectrum was acquired from this thesis. The signals were assigned according to the previously reported chemical shifts of M1-diUb (BMRB entry: 26790). Signals of the sidechains are linked via dashed lines and not assigned. NMR signals were assigned by CCPNMR. Figure was generated by NMRviewJ.

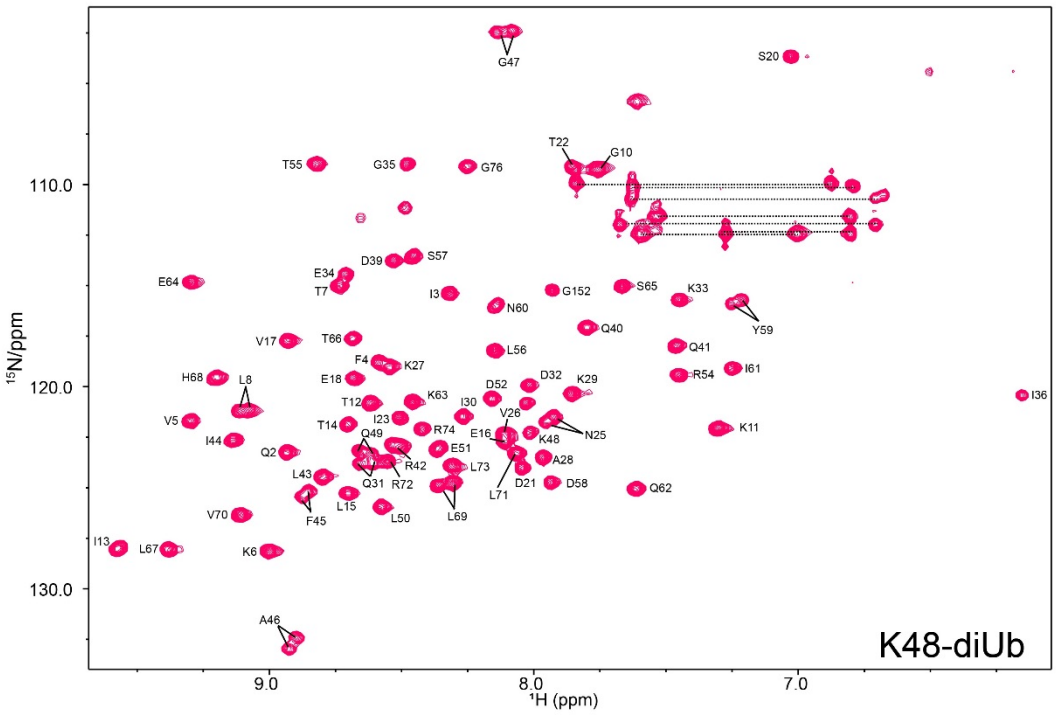


Figure A2 Full ^{1}H/^{15}N-HSQC assignments of K48-diUb. The spectrum was acquired from this thesis. The signals were assigned according to the previously reported chemical shifts of ubiquitin (BMRB entry: 17769) and M1-diUb (BMRB entry: 26790). Signals of the sidechains are linked via dashed lines and not assigned. NMR signals were assigned by CCPNMR. Figure was generated by NMRviewJ.

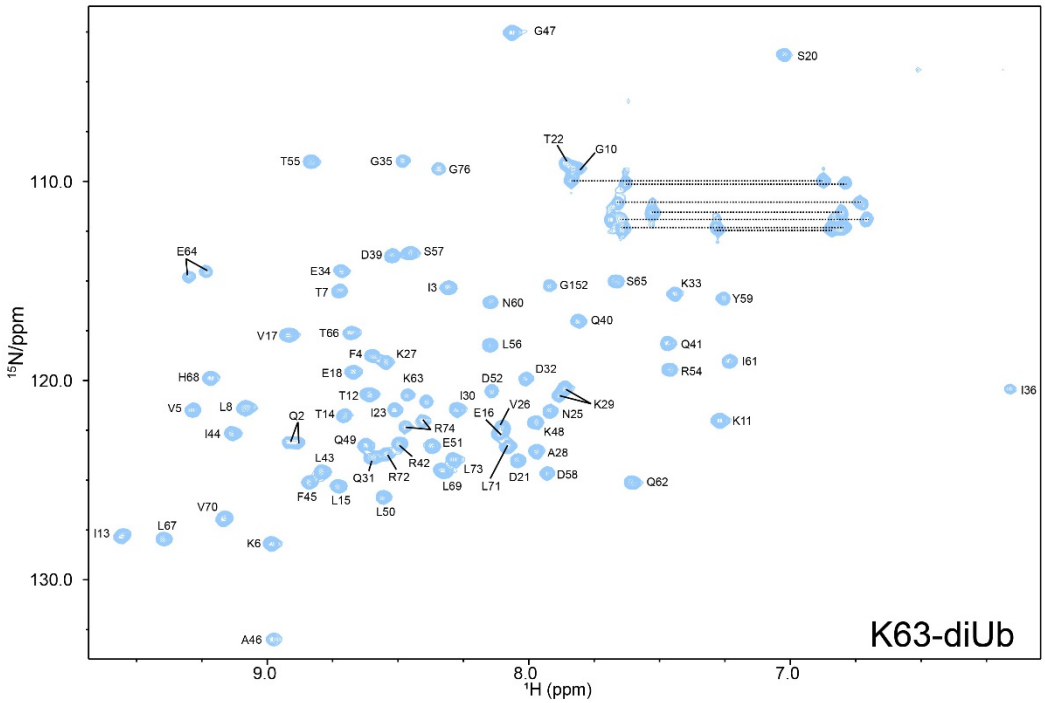


Figure A3 Full ^{1}H/^{15}N-HSQC assignments of K63-diUb. The spectrum was acquired from this thesis. The signals were assigned according to the previously reported chemical shifts of ubiquitin (BMRB entry: 17769) and M1-diUb (BMRB entry: 26790). Signals of the sidechains are linked via dashed lines and not assigned. NMR signals were assigned by CCPNMR. Figure was generated by NMRviewJ.

Appendix III

Sequence alignments of N4BP1 with MCPIP1 and KHNYN

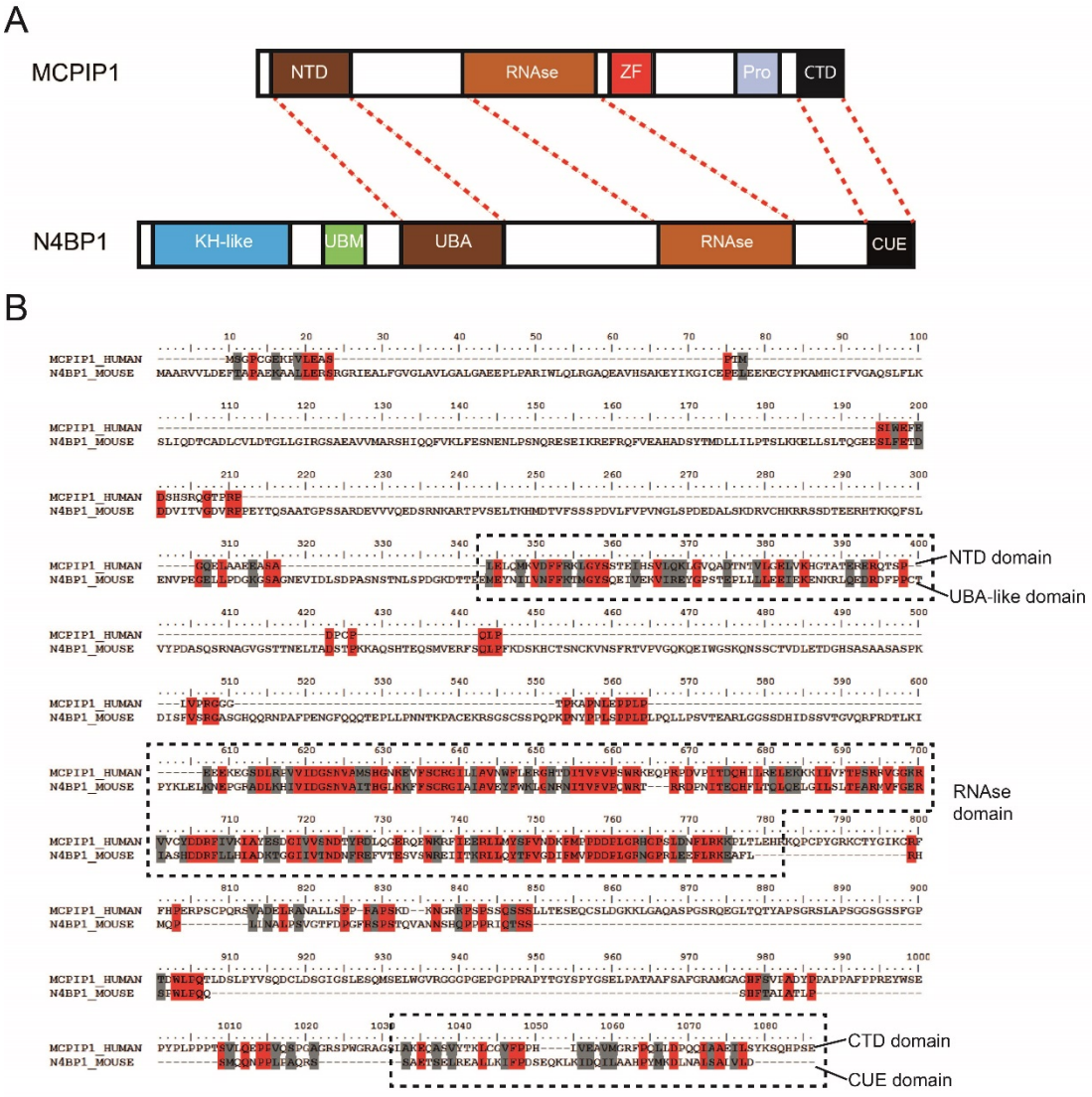


Figure A4 Sequence comparison between mN4BP1 (Uniprot: Q6A037) and hMCPIP1 (Uniprot: Q5D1E8). Domains sharing high sequence similarity were highlighted in dashed boxes. A. Schematic representation of the domain structures from the two proteins. B. Full-length sequence alignments between the two proteins. Identical residues are shaded in red. The residues with >80% similarities are shaded in grey. Sequences were aligned by MEGA and visualized by bioedit.

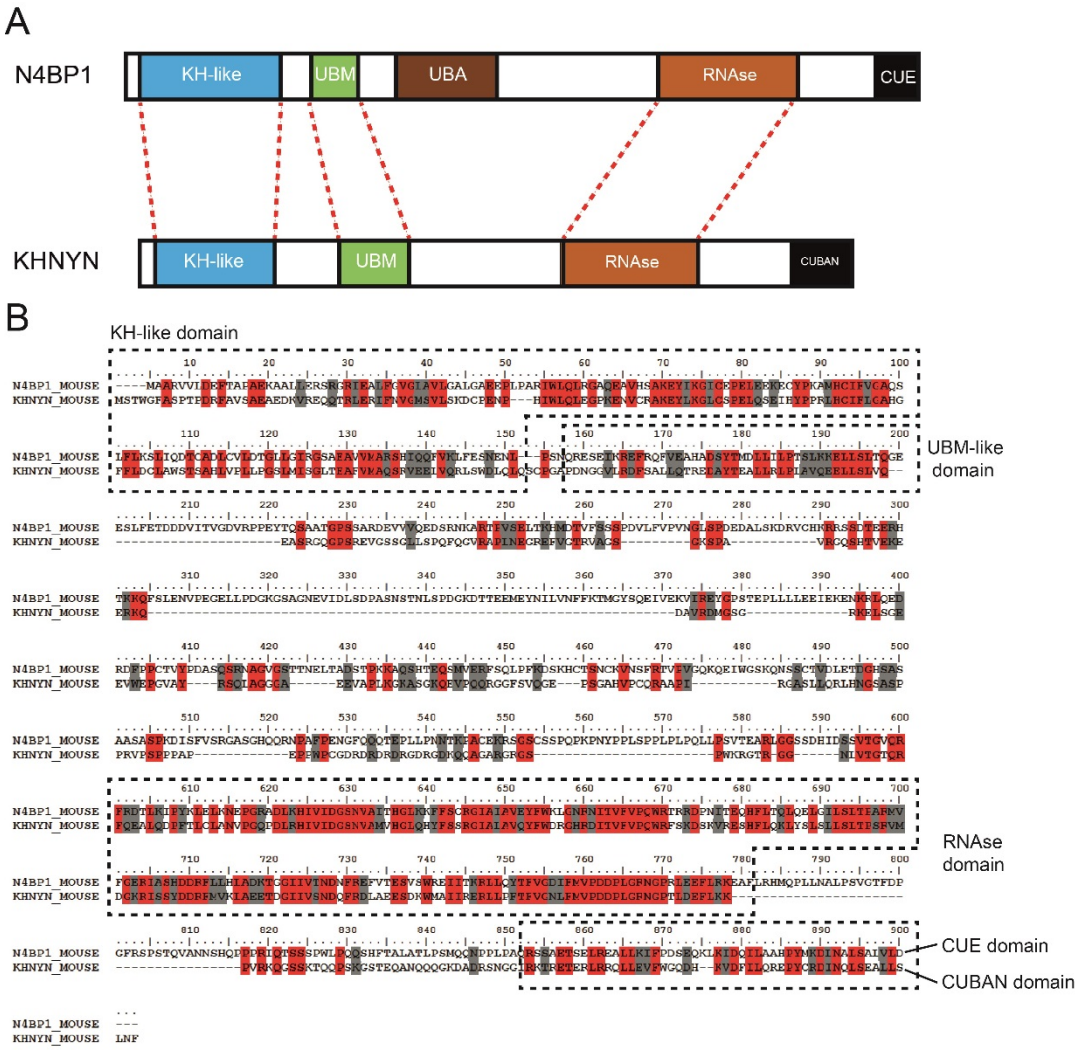


Figure A5 Sequence comparison between mN4BP1 (Uniprot: Q6A037) and mKHNYN (Uniprot: Q80U38). N4BP1 shares overall high sequence similarity with KHNYN except that KHNYN does not contain the UBA-like domain. Domains sharing high sequence similarity were highlighted in dashed boxes. A. Schematic representation of the domain structures from the two proteins. B. Full-length sequence alignments between the two proteins. Identical residues are shaded in red. The residues with >80% similarities are shaded in grey. Sequences were aligned by MEGA and visualized by bioedit.

Appendix IV

Protein constructs and primers used in this thesis

Table A1 N4BP1 constructs used in this thesis

Construct Name	Domain description	Vector	Forward primer	Reverse primer
mN4BP1 (1-392)	KH+UBM+UBA	pGEX-6P1	N4BP1-F	N4BP1-392R
mN4BP1 (1-342)	KH+UBM	pGEX-6P1	N4BP1-F	N4BP1-342R
mN4BP1 (1-311)	KH+UBM	pGEX-6P1	N4BP1-F	N4BP1-311R
mN4BP1 (1-144)	KH-like domain	pET47b	47b-N4BP1-F	47b-N4BP1-144-R
mN4BP1 (144-311)	UBM-like domain	pET47b	47b-N4BP1-144-F	47b-N4BP1-311-R
mN4BP1 (312-392)	UBA-like	pET47b	47b-N4BP1-312-F	47b-N4BP1-392-R
mN4BP1 (613-774)	RNAse	pCold-TF	pCold-RNAse-F	pCold-RNAse-R
hMCPIP1 (112-132) + mN4BP1 (613-774)	Chimeric RNAse	pCold-I	pCold-NM; NM-1; NM-2	pCold-RNAse-R
mN4BP1 (613-893)	RNAseCUE	pCold-TF	pCold-RNAse-F	pCold-CUE-R
hMCPIP1 (112-132) + mN4BP1 (613-893)	Chimeric RNAseCUE	pCold-I	pCold-NM; NM-1; NM-2	pCold-CUE-R
mN4BP1 (850-893)	CUE	pET47b	47b-CUE-F	47b-CUE-R
mN4BP1 (850-893)	CUE	pGEX-6P1	6P1-CUE-F	6P1-CUE-R
mN4BP1 (850-893)	CUE (F862G/P863A)	pGEX-6P1	FP-GA-F	FP-GA-R

Table A2 Other protein constructs used in this thesis

Construct Name	Domain description	Vector	Note
hHOIP (697-1072)	RBR+LDD	pET49b	Lab stock
hHOIP (300-1072)	ZF+NZF1+NZF2+UBA+RBR+LDD	pET49b	Lab stock
hHOIL-1L (1-478)	FL protein lacking the last 32 residues	pET49b	Lab stock
CDC34		pET15	Lab stock
M1-diUb		pET47b	Lab stock
M1-tetraUb		pGEX-4T1	Lab stock
M-Cys-Ub		pGEX-6P1	Lab stock
NEDD8		pGEX-2TK	Lab stock
Ub WT		pGEX-6P1	Lab stock
Ub K48A		pGEX-6P1	Lab stock
Ub K48R		pGEX-6P1	Lab stock
UbcH5c		pGEX-6P1	Lab stock
UBC13		pET15b	Lab stock
Ube1		pET28	Lab stock
UBE2V1 (UEV1a)		pET15b	Lab stock

Table A3 Primers used for making the N4BP1 protein constructs

Primer Name	Sequence (5'-3')
47b-N4BP1-F	CTTTCAGGGACCCGGGATGGCGGCCCGGGTG
47b-N4BP1-144-F	CTTTCAGGGACCCGGGAACGAGAACCTCCCCAGTAATCAG
47b-N4BP1-312-F	CTTTCAGGGACCCGGGGGTAAGGGGTCAGCTGGAAATG
47b-N4BP1-144-R	TTAATTAAGCCTCGAGTCAGTTGCTCTCAAAGAGCTTCACAAAC
47b-N4BP1-311-R	TTAATTAAGCCTCGAGTCAATCTGGCAAAAGCTCCCCTTC
47b-N4BP1-392-R	TTAATTAAGCCTCGAGTCATTCTTGGAGCCTTTTATTTTCTTTTTC
47b-CUE-F	CTTTCAGGGACCCGGGGAGACCAGCGAGCTGAGGG
47b-CUE-R	TTAATTAAGCCTCGAG TCAGTCCAACACCAGGGCC
6P1-CUE-F	GGGGCCCCTGGGATCCGAGACCAGCGAGCTGAGGG
6P1-CUE-R	TCACGATGCGGCCGCTCAGTCCAACACCAGGGCC
FP-GA-F	GAGGCGCTGCTAAAGATCGGCGCTGACTCTGAGCAAAAGC
FP-GA-R	GCTTTTGCTCAGAGTCAGCGCCGATCTTTAGCAGCGCCTC
NM-1	GGTGGTGGCACCCCTAAGGCTCCCAACCTGGAGCCTCCACTCCCAGAAG
NM-2	CCTCCACTCCCAGAAGAGGAAAAGGAGGGCGCAGATTTGAAGCAT
N4BP1-F	GGGGCCCCTGGGATCCATGGCGGCCCGGGTG
N4BP1-392R	TCACGATGCGGCCGCTCATTCTTGGAGCCTTTTATTTTCTTTTTC
N4BP1-342R	TCACGATGCGGCCGCTCATTCTGTAGTGTCTTTTCCATCGGG
N4BP1-311R	TCACGATGCGGCCGCTCAATCTGGCAAAAGCTCCCCTTC
pCold-NM	GAAGGTAGGCATATGGGTGGTGGCACCCCT
pCold-RNAse-F	GAAGGTAGGCATATGGCAGATTTGAAGCATATTGTTATAGATGG
pCold-RNAse-R	ACCGAGCTCCATATGTCAAAAGGCCTCCTTTCGAAGAAAT

Appendix V

Structure statistics for N4BP1 CUE CS-Rosetta model

Table A4: Summary of structure statistics for N4BP1 CUE CS-Rosetta model

Experimental restraints input for CS-Rosetta	
$^{13}C^{\alpha}$ shifts	44
$^{13}C^{\beta}$ shifts	44
$^{13}C'$ shifts	43
^{15}N shifts	40
$^{1}H^{N}$ shifts	40
$^{13}H^{\alpha}$ shifts	44
Total restraints	255
Average pairwise RMSD* (Å)	
C^{α}	0.49
Backbone atoms	0.48
Heavy atoms	0.80
All atoms	1.01
Measures of structure quality**	
Ramachandran favored regions (%)	100
Ramachandran Outliers (%)	0
Ramachandran distribution Z-score	
whole	-1.43 ± 0.90
helix	-0.69 ±0.60
sheet	none
loop	-1.64 ± 1.39
Favored rotamers (%)	100
Poor rotamers (%)	0
C^{β} deviations >0.25 Å	0
Bad bonds (%)	0
Bad angles (%)	0
All-atom clashscore***	1.4

* Pairwise RMSD was calculated using the structure ensemble containing 10 best refined models.

**Quality data for model 5 from the Structure ensemble. Model 5 is the overall representative, medoid model (most similar to other models in the ensemble).

***Clashscore is the number of serious steric overlaps (>0.4 Å) per 1000 atoms.

Appendix VI

Supplementary binding data

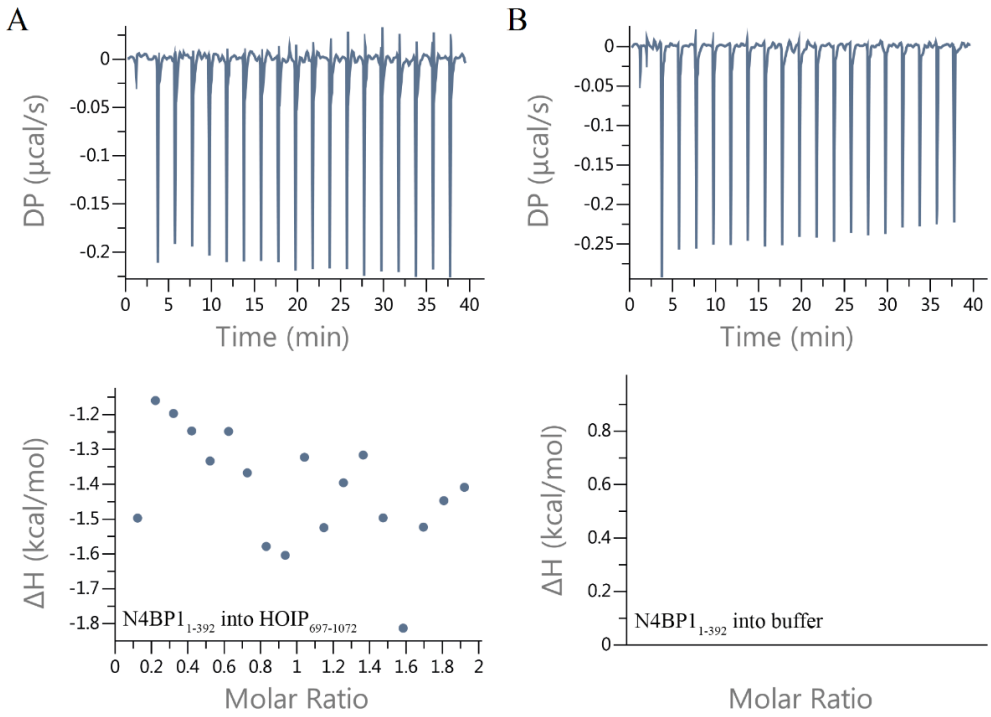


Figure A6 ITC measurement reveals that N4BP1$_{1-392}$ doesn't bind to HOIP$_{697-1072}$. A: The isotherm of titrating N4BP1$_{1-392}$ into HOIP$_{697-1072}$. B: The isotherm of titrating N4BP1$_{1-392}$ into buffer as control. Data were analyzed by the MicroCal PEAQ-ITC Analysis Software. Measurement conditions: Syringe: 0.5 mM N4BP1$_{1-392}$, Cell: 50 µM HOIP$_{697-1072}$, 18 injections, 2 µL/injection, duration: 2s, spacing time: 120 s, measurement temperature: 25°C.

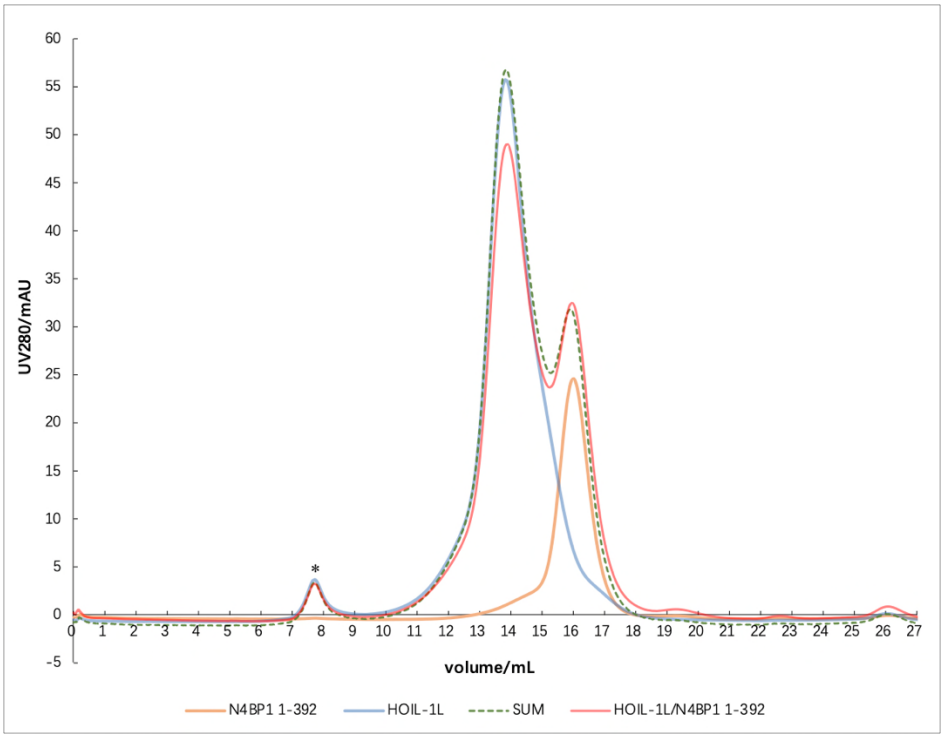


Figure A7 Analytical size-exclusion chromatography reveals that N4BP1$_{1-392}$ doesn't bind to HOIL-1L. The SEC column used here is Superose 10/300 GL (24 mL, GE Healthcare). The buffer used here is SEC buffer (150 mM NaCl, 50 mM HEPES, pH 7.6, 1 mM DTT). Flow rate: 0.5 mL/min. Each protein in the sample is 100 μM. Loading volume is 200 μL for all the measurements. "HOIL-1L/N4BP1 1-392" indicates the UV trace of the mixture of N4BP1$_{1-392}$ and HOIL-1L (each at 100 μM) together loaded onto the column. No peak shift was observed for the protein mixture indicating that there was no larger protein complex formed. "SUM" is the sum of the two separate UV traces of N4BP1$_{1-392}$ and HOIL-1L. HOIL-1L: 54.5KDa. N4BP1(1-392): 34KDa. * indicates the impurity from HOIL-1L.

Appendix VII

Investigating the role of an arginine residue near the Fe-S cluster N2 in mitochondrial complex I from *Yarrowia lipolytica*

Supervisor: Dr. Maxie Roessler

Table of contents

1.1 Introduction

Mitochondria are among the most important and complicated organelles in eukaryotes. Originated from endosymbiosis, mitochondria are semiautonomous and involved in various processes, including energy metabolism, signal transduction, redox homeostasis etc. The main function of mitochondria is supplying energy to the cells. During aerobic respiration, products from glycolysis, β-oxidation of fatty acid as well as amino acid oxidation are transferred to the tricarboxylic acid cycle (TCA), and the energy is reserved in the form of two reductive substrates—NADH and $FADH_2$. In mitochondria, electrons from the reductive substrates are transferred through the electron transport chain (ETC) and then reduce O_2 to finally generate H_2O. At the same time, the protons are pumped from the mitochondria matrix to the intermembrane space, generating the so-called "proton-motive force (Δp)". Protons then follow the electrochemical gradient to return to the matrix via ATP synthase. ATP is thus generated.

The electron transport chain consists of a series of redox carriers, including NAD^+/NADH, flavoprotein, cytochrome, iron-sulfur protein and ubiquinone. These basic units are further assembled into four complexes (complex I-IV) (Table B1; Figure B1).

Table B1 The composition of mammalian electron transport chain

Complex	Full name	Mw	Redox centers
I	NADH: ubiquinone oxidoreductase	980 KDa	FMN, two [2Fe-2S], six [4Fe-4S]
II	succinate dehydrogenase	120 KDa	FAD, [2Fe-2S], [4Fe-4S], [3Fe-4S], cyt b_{560}
III	cytochrome bc_1 complex	480 KDa	cyt b_{562}, cyt b_{566}, Rieske FeS, cyt c_1
IV	cytochrome c oxidase	420 KDa	Cu_A, cyt a, Cu_B/cyt a_3

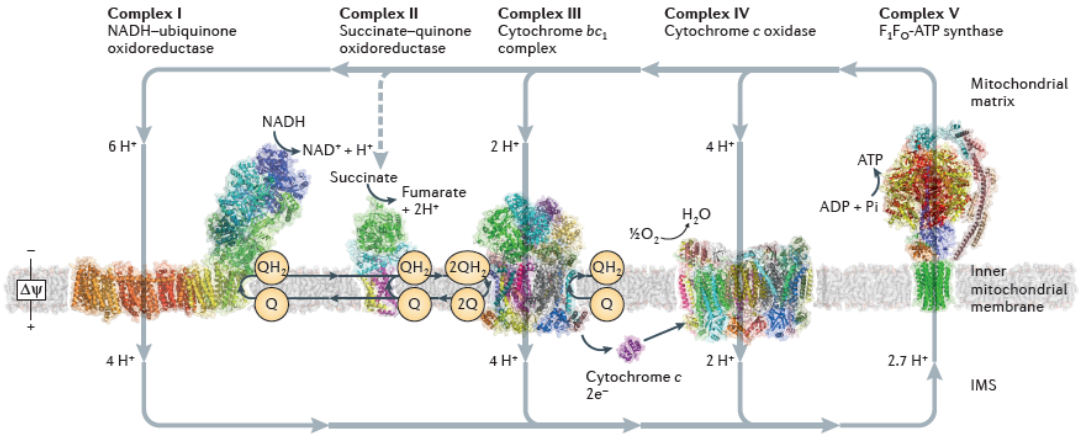


Figure B1. The mitochondrial respiratory chain showing all four complexes and ATP synthase (© 2015 Sazanov, by permission, Licence No: 4872620687820).

Complex I (NADH: ubiquinone oxidoreductase) is the largest protein complex in the respiratory chain, including one flavin mononucleotide and eight iron-sulfur clusters. Complex II (succinate dehydrogenase) has four subunits and six cofactors. The electrons are transferred through the iron-sulfur clusters from both Complex I and Complex II, and finally reach to the ubiquinone binding site to reduce ubiquinones (UQ or CoQ). Complex III (cytochrome bc_1 complex) catalyzes the electron transfer from reduced ubiquinone (UQH_2) to cytochrome c. Complex IV (cytochrome c oxidase) is the last enzyme in the respiratory chain. The electrons are transferred from cytochrome c to reduce O_2 into H_2O.

For each pair of electrons transferred to O_2 through the respiratory chain, four protons are pumped out through the inner mitochondrial membrane by Complex I, four by Complex III and two by Complex IV. The proton gradient thus forms the proton-motive force to generate ATP through ATP synthase (Figure B1). The overall equation for this process is:

$$NADH + 11H^+ + 1/2\ O_2 \rightarrow NAD^+ + 10H^+ + H_2O$$

1.1.1 Complex I

Complex I (NADH: ubiquinone oxidoreductase), also known as NADH dehydrogenase, at a molecular weight of ~980 KDa in mammalian species, is one of the largest membrane protein assemblies and one of the most important components in the mitochondria respiration chain.

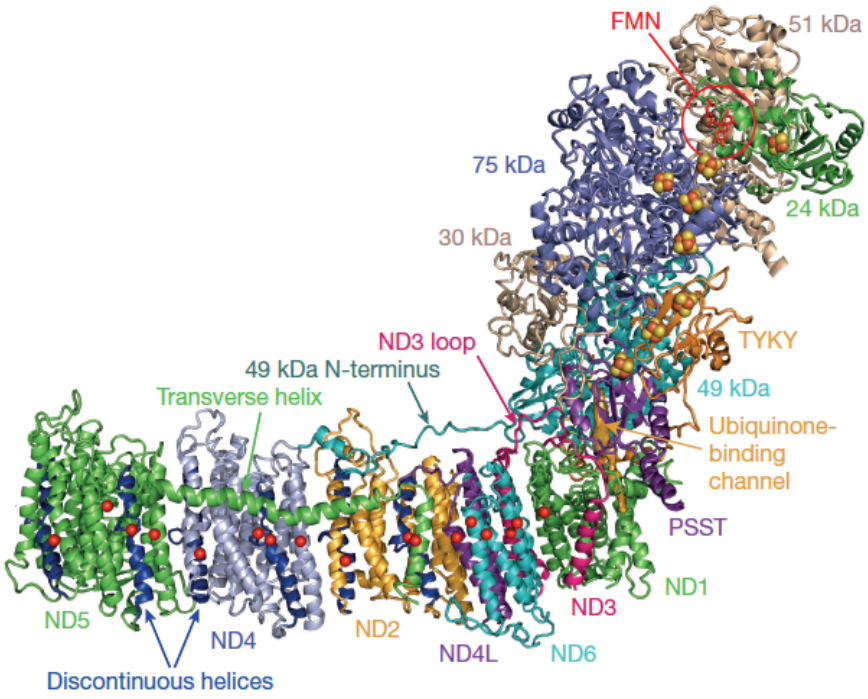


Figure B2 A brief structure of the core subunits of mammalian Complex I showing the peripheral arm and the membrane arm (© 2016 Zhu et al., by permission, Licence No: 4872630418512).

Electron microscopic reconstructions have established the overall L-shaped structure of Complex I from all reported organisms (Baradaran et al., 2013; Djafarzadeh et al., 2000; Grigorieff, 1998). The peripheral arm protrudes into the mitochondrial matrix and the membrane arm is embedded in the inner membrane. Mammalian Complex I consists of 45 subunits, of which 14 make up the catalytic core (Figure B2). The core subunits are highly conserved from bacteria to human, dominating the catalytic machinery (Zhu et al., 2016). Because the main research object in this chapter is the Complex I

from *Yarrowia lipolytica* (a yeast species), the corresponding names of the core subunits from *Bos taurus* and *Yarrowia lipolytica* are compared in Table B2.

Table B2 The nomenclature of 14 core subunits of Complex I

Domain	*Bos taurus*	*Yarrowia lipolytica*	Cofactors
Nuclear genome encoding, hydrophilic arm	75 KDa	NUAM	one [2Fe-2S], two [4Fe-4S]
	51 KDa	NUBM	flavin, one [4Fe-4S]
	49 KDa	NUCM	
	30 KDa	NUGM	
	24 KDa	NUHM	
	PSST	NUKM	one [2Fe-2S]
	TYKY	NUIM	one [4Fe-4S] (cluster N2)
Mitochondrial genome encoding, hydrophobic arm	ND1	NU1M	
	ND2	NU2M	
	ND3	NU3M	
	ND4	NU4M	
	ND5	NU5M	
	ND6	NU6M	
	ND4L	NULM	

Complex I first oxidizes one molecule of NADH, then delivers two electrons through a long chain of iron-sulfur clusters and finally reduces one molecule of ubiquinone. Concomitantly, four protons are pumped into the mitochondrial matrix, thus contributing ~ 40% of the proton-motive force to the ATP synthesis. These processes are described as proton-coupled electron transfer (PCET) reactions. The overall equation of this catalysis is:

$$NADH+H^{+}+UQ+4H^{+}_{in} \rightarrow NAD^{+}+UQH_2+4H^{+}_{out}$$

As a major contributor of reactive oxygen species (ROS), Complex I dysfunction has been implicated in various diseases such as Leigh disease, Leber's hereditary optic neuropathy etc. Moreover, Complex I is the most common site for mitochondrial anomalies, accounting for as many as one-third of respiratory chain deficiencies. Genetic engineering by mouse model has revealed several important roles of different Complex I subunits. Understanding how Complex I functions at the molecular level will ultimately assist to develop powerful medical treatments to the mitochondria-related diseases.

In Complex I, NADH oxidation and electron transfer have been now relatively well understood. In contrast, the mechanism of how electron transfer and ubiquinone reduction are involved in proton translocation is still ambiguous (Hirst & Roessler, 2016).

1.1.2 Iron-sulfur clusters

Iron-sulfur clusters are found in metalloproteins including ferredoxins, hydrogenases, and mitochondrial Complex I, acting as key cofactors in the redox reaction to transfer electrons. For over 40 years, electron paramagnetic resonance spectroscopy (EPR) has been employed as a powerful way to characterize Fe-S clusters. There are 8 iron-sulfur clusters in mammalian Complex I, seven of which participate in the process of electron transfer. Among them, cluster N1a and N1b are binuclear [2Fe-2S] clusters, while cluster N2, N3, N4, N5, N6a and N6b are tetranuclear [4Fe-4S] clusters. Bacteria Complex I contains an additional [4Fe-4S] cluster, named N7, but not conserved. When Complex I is reduced with NADH, some of the clusters are paramagnetic thus can be detected by EPR spectroscopy (for example, N1b, N2, N3, N4 as shown in Figure B3).

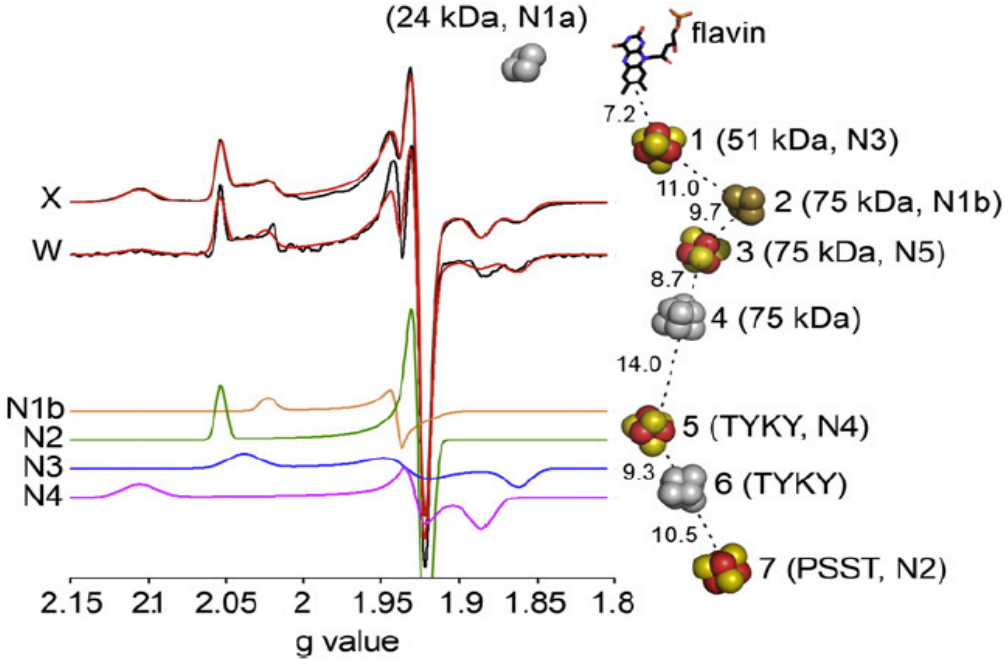


Figure B3 Fe-S clusters and the electron transfer relay in Complex I. The X-band and W-band EPR spectra showing the overall signals of Fe-S clusters in Complex I from *B. taurus*. Spectra coloured in black are the original data while spectra coloured in red are the simulated data. The subspectra of four Fe-S clusters are isolated by simulation and indicated respectively below the X-band and W-band EPR spectra. The electron transfer relay is shown next to the spectra. Clusters coloured in grey don't contribute to EPR signals (© 2016 Hirst and Roessler, by permission).

Fe-S clusters in Complex I are arranged in a specific relay to transfer electrons, thus defining the main electron pathway (Hirst & Roessler, 2016). Figure B3 shows the typical EPR spectra of NADH reduced Complex I from *B. taurus*, of which four Fe-S clusters can be identified. N1 signal from *Y. lipolytica* has been shown to correspond to the N1b signal from *B. taurus* (Zickermann et al., 2007). Note that only at a temperature below 10 K and higher microwave power, can cluster N5 be detectable due to its fastest spin relaxation rate (Kerscher et al., 2002). In the order of increasing spin relaxation rate, the sequence of these clusters is N1, N2, N3, N4 and N5.

1.1.3 Cluster N2

Cluster N2 is in the terminal of the electron transfer pathway in Complex I and localized in the interface between the PSST and 49 KDa subunit in *Bos taurus*. While all the other EPR-detectable clusters in this pathway show a relatively lower redox midpoint potential, N2 exhibits the highest midpoint potential (the $E_{m,7}$ of N2 is ~-140 mV). It receives electrons from upstream Fe-S clusters and then transfers to ubiquinone (Ohnishi, 1998).

1.1.3.1 The study of cluster N2 by EPR spectroscopy in *Y. lipolytica*

The pH dependence of cluster N2 and its close distance to the ubiquinone binding site has made it attractive to investigate the role of N2 in proton translocation (Zwicker et al., 2006). Most of the research on N2 is conducted by EPR spectroscopy. Thanks to the genetic tools using *Y. lipolytica,* site-directed mutagenesis of the nuclear genome encoded subunit can be easily achieved. Thus, several critical residues that are close to the N2 cluster have been mutated to study the coupled mechanism. Table B4 shows the reported changes of the N2 EPR signal from corresponding point mutation.

From the EPR spectra, several key residues that display significant changes of EPR signals have been identified. Shifted EPR signal may indicate a possible change of the microenvironment around N2. The decreased or increased intensity of N2 signals may imply the changes of the N2 potential. Most of the reported EPR spectra of N2 cluster are in relatively low resolution or limited by the sample preparation (some mutations were detected using the membrane prep). It still could not give a whole picture of how electron transfer in cluster N2 is coupled with proton translocation. To this end, advanced EPR methods with higher resolution may bring this to light.

Table B3 Reported N2 EPR signals of mutations in NUCM subunit (Bovine: 49 KDa) from *Y. lipolytica*

Mutant	g_z signal of N2	Sample preparation	EPR condition	Reference
H120A	normal	Complex I reduced by NADH	12K, 9.48GHz, 2mW, 0.8mT	Kashani-Poor et al., 2001
R141A	undetectable			
D143E	decreased			
Y144H	shifted to higher value			
V460A	undetectable			
V460M	decreased			
E463Q	decreased			
H226A	undetectable		12K, 9.48GHz, 1mW, 0.64mT	Grgic et al., 2004
H226Q	<10%			
H226C	<10%			
H226M	reduced by ~2/3, shifted			
R141A	undetectable	membrane containing Complex I, reduced by NADH	12K, 9.48GHz, 2mW, 0.8mT	
R141K	5%			
R141M	undetectable			
Y144F	normal		12K, 5mW	Tocilescu et al., 2010
Y144W	g_{xy} shifted to lower field			
Y144I	undetectable			
Y144H	slightly shifted to higher value			
Y144S	decreased			
Y144R	undetectable			

1.1.3.2 Possible participation of cluster N2 in proton translocation

Some evidence has partially supported that cluster N2 is involved in coupling redox energy to proton translocation. The highly conserved residue Tyr87, located in the Nqo4 subunit of Complex I in *Thermus thermophilus*, close to cluster N2 and facing the quinone-binding site, has been suggested to participate in both electron and proton transfer (Berrisford & Sazanov, 2009). Cluster N2 has an unusual coordination by 'tandem cysteines', Cys54 and Cys55 (Figure B4), which has been proposed to initiate conformational changes and/or protonation of bound quinone (Berrisford & Sazanov, 2009).

The reduction potential of cluster N2 is pH-dependent at a value of -60 mV/ pH within physiological pH range and it has been demonstrated that a -60 mV/ pH is needed for a strong proton-coupled electron transfer reaction (Ingledew & Ohnishi, 1980), giving cluster N2 much possibility to act as an energy-coupling site. However, the reported pH dependence value of cluster N2 is -36 mV/ pH in *Y. lipolytica* (Zwicker et al., 2006). Through site-directed mutagenesis in *Y. lipolytica*, the conserved residue His226 has been identified as the major redox-Bohr group that modulates the N2 redox potential. At pH 7, the midpoint potential of cluster N2 in H226M shifted from -140 mV to -220 mV and the pH dependence of cluster N2 was completely abolished (Grgic et al., 2004; Zwicker et al., 2006). Recently, Le Breton and co-workers re-measured the pH dependence of cluster N2 and concluded that cluster N2 from bovine Complex I possessed a pH dependence of -40 mV per pH unit which is in agreement with the observation from *Y. lipolytica*. Advanced EPR study of the micro-environment of cluster N2 has demonstrated that two weak-coupled exchangeable protons are present near cluster N2 but not strong enough to

be coupled with the energy transduction, thus excluding the direct role of cluster N2 as the energy-coupling site in Complex I (Le Breton et al., 2017).

Even so, the possible indirect role of cluster N2 in Complex I catalysis cannot be excluded and its clear function requires further investigation.

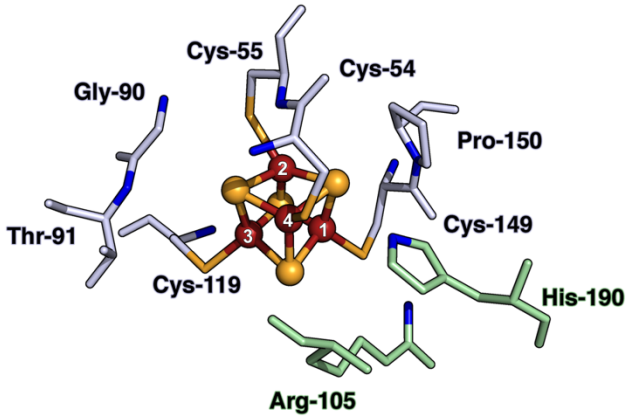


Figure B4 Conserved key residues around cluster N2 with bovine complex I numbering (© 2018 J. Wright, by permission, private correspondence).

1.1.4 Quinone reduction

In Complex I, ubiquinone is reduced by its direct electron donor cluster N2. Ubiquinone accepts one electron to become semiquinone radicals, which further receive one electron to form ubiquinol (QH_2). It is quite free for ubiquinone to diffuse in the inner mitochondrial membrane because of its hydrophobic property, making itself a perfect carrier to transfer electrons to Complex III in the respiratory chain. Notably, ubiquinone is competent to carry both electrons and protons, thus playing a critical role in coupling redox energy to proton translocation.

The 3.3 Å crystal structure of the entire Complex I from *T. thermophilus* has provided several important clues to the mechanism of quinone binding and proton translocation. Co-crystallization of this bacterial Complex I with decylubiquinone reveals that the quinone headgroup is located about 12 Å from cluster N2 to guarantee the electron transfer, and forms hydrogen bond

with H38 and Y87 of the subunit Nqo4 (corresponding to the 49KDa in bovine), thus a quinone-binding channel was proposed (Baradaran et al., 2013). By hybrid quantum mechanics and molecular mechanics simulations in *T. thermophilus*, a hypothesis was proposed that the quinone reduction is coupled to its local protonation by three residues from Nqo4. Tyr87 and His38 rapidly transfer one proton respectively to quinone to generate QH_2. The local deprotonation of Asp139 by His38 further stabilized the QH_2 species. The His-Asp forms the ion pair that undergoes a large conformational change during this process. When quinone receives an electron from the cluster N2, the His-Asp pair remains intact. But after the protonation of quinone, the two residues from this ion pair dissociate from each other. Meanwhile, the hydrogen bond between Y87 and the headgroup of QH_2 is destroyed (Sharma et al., 2015). From the structure of mammalian Complex I with a 4.2 Å resolution, a similar quinone-binding channel is identified, indicating the high conservation between prokaryotes and eukaryotes (Zhu et al., 2016). The conserved critical residues that form the hydrogen bond with quinone can be identified as well (Figure B5 left). Interestingly, a "bottleneck" composed by several charged and polar residues was described in this channel, suggesting that conformational changes may occur to ensure appropriate quinone binding and dissociation (Figure B5 right).

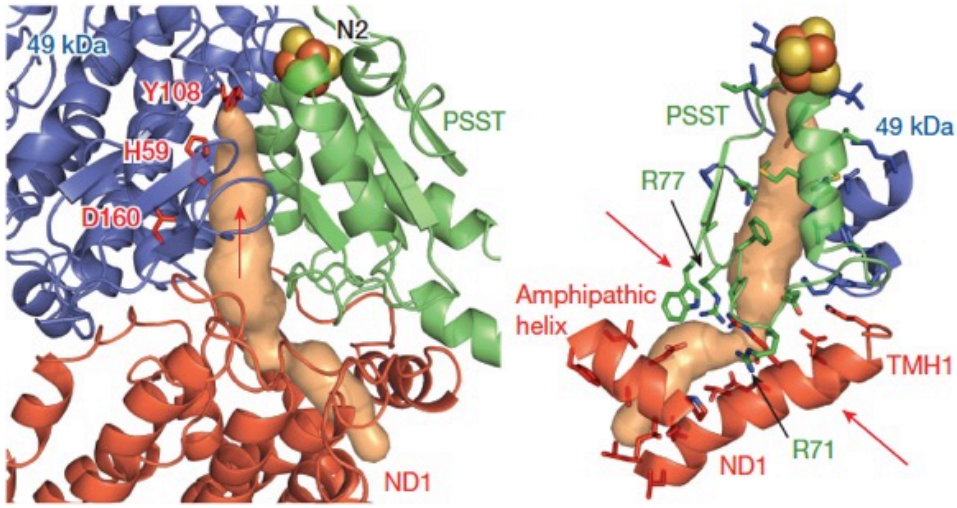


Figure B5 The proposed quinone-binding channel in mammalian Complex I from bovine (© 2016 Zhu et al., by permission, Licence No: 4872630418512)**.** The bottleneck is indicated by the two red arrows on the right.

1.1.5 Protein arginine methylation

Protein arginine methylation is one of the most prevalent and conserved post-translational modifications. It is widely spread in eukaryotes and catalyzed by a group of protein family called protein arginine methyltransferases (PRMTs). PRMTs could transfer the methyl group from S-adenosylmethionine (AdoMet) to the guanidine group of the arginine residue in the target protein.

Figure A6 shows the methylation pathway catalyzed by PRMTs (Esse et al., 2012). In general, PRMTs can be classified into two main subfamilies, both of which could catalyze mono-methylation process to produce mono-methylarginine (MMA). The difference between the two families is that Type I RPMTs function to further catalyze asymmetric dimethylation where both two methyl groups are covalently attached to the same nitrogen atom of the guanidine group. This family includes PRMT1, PRMT3, PRMT4, PRMT6 and PRMT8, producing asymmetric dimethylarginine (ADMA). Whereas Type II RPMTs catalyze symmetric dimethylation to generate symmetric

di-methylarginine (SDMA) where two methyl groups are attached to two different nitrogen atoms of the guanidine group. This family contains PRMT5 and PRMT9 (Bedford, 2007; Yang & Bedford, 2013).

It is found that nearly 2% of the arginine residues are dimethylated in the liver cell nucleus. This modification has been implicated to participate in various critical biological processes, including transcriptional regulation, RNA metabolism, chromatin remodelling, signaling transduction and cell proliferation etc (Bedford & Clarke, 2009; Boffa et al., 1977).

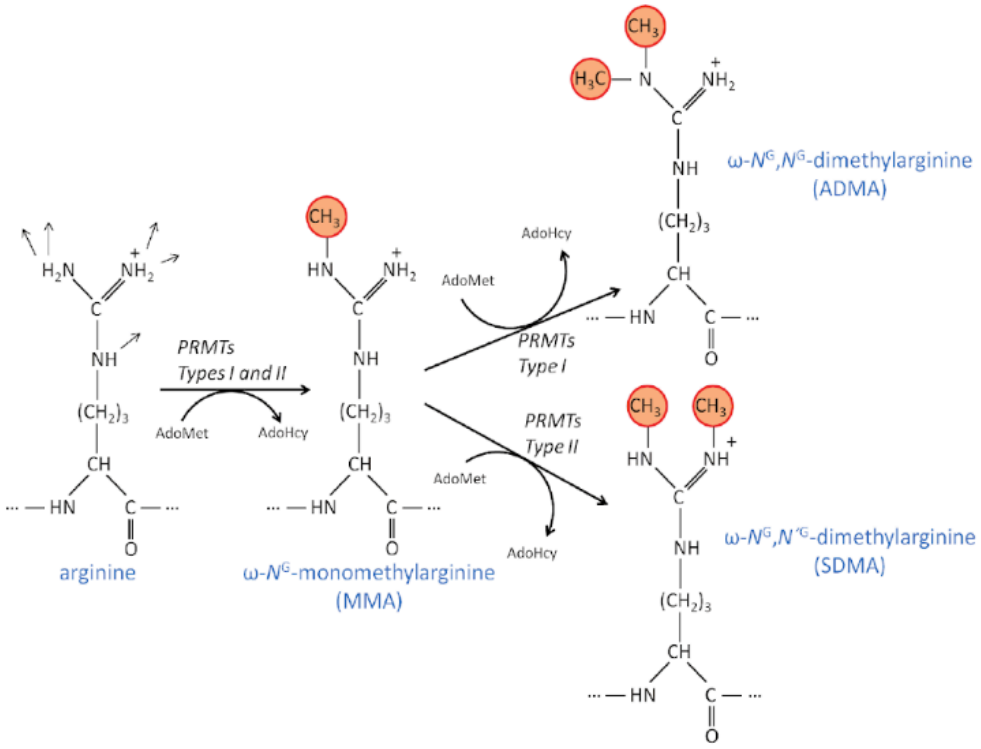


Figure B6 The dimethylation pathway catalyzed by PRMTs (© 2012 Esse et al., by permission).

1.1.5.1 Dimethylation of Arginine 121 in the Complex I NUCM subunit

Residue Arg121 is located in the NUCM subunit of Complex I from *Y. lipolytica*. It became the first candidate residue for investigation in this

project. Some of its distinctive characteristics make it attractive to study the effect on the reduction potential of cluster N2.

First of all, R121 is in close proximity to cluster N2 (~5 Å) (Figure B7a), and well-conserved among several species from *E. coli* to human, but surprisingly absent in *T. thermophilus* (Figure B7c). Another special feature of this residue is that it is dimethylated in human, bovine, yeast and some of the bacteria but not in *E.coli* (Figure B7b). As for the fact that the N2 potential is both much lower in *T. thermophilus* and *E.coli* than other species, Hirst and Roessler (2016) suggested that the presence of this modification on the arginine residue influences the reduction potential of cluster N2.

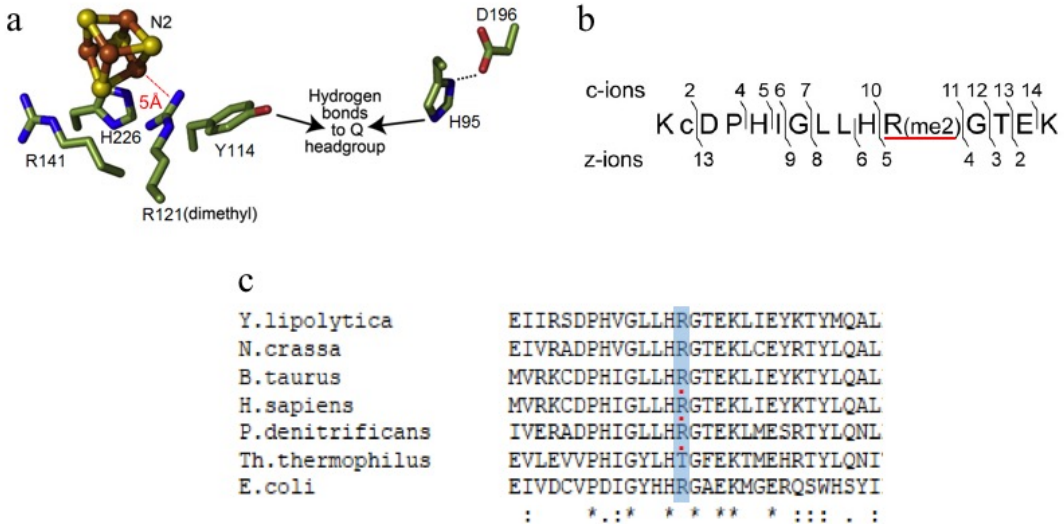


Figure B7 The arginine residue in the NUCM subunit of Complex I. a. The position of R121 residue (in *Y. lipolytica* Complex I numbering) relative to N2. Figure modified from the bovine Complex I structure (© 2016 Hirst and Roessler, adapted with permission); b. The evidence from mass spectroscopy that R85 in the human 49KDa subunit is dimethylated as indicated by the red line (© 2013 Carroll et al., by permission); c. Sequence alignments showing the conservation of the arginine residue among several species. the red dots indicating that the residue is demethylated in the corresponding species.

1.1.5.2 The function of arginine dimethylation in Complex I

There are some functional studies on the R121 residue from different spices. J. E. Walker's group published two papers in 2013 on the topic of R85 dimethylation (the corresponding residue of R121 in human). One shows that R85 can be symmetrically dimethylated at its ω–N^{G}, $N^{G'}$ atoms of the guanidine group (Carroll et al., 2013). In another paper, it is confirmed that NDUFAF7 is the methyltransferase which is responsible for dimethylating the R85 residue in the human NDUFS2 subunit (corresponding to the 49KDa in bovine Complex I) (Rhein et al., 2013). NDUFAF7 belongs to the type II PRMTs. Suppression of NDUFAF7 causes the severe decrease of OCR (rate of consumption of oxygen) linked to Complex I. Moreover, the expression of both NDUFS7 (corresponding to PSST in bovine Complex I) and ND1 subunit is reduced, indicating that NDUFAF7 is required for the assembly of Complex I. Consistent with this, MidA, the homologous protein of human NDUFAF7 in *Dictyostelium discoideum*, has been identified as the methyltransferase required for the assembly of mitochondrial Complex I (Carilla-Latorre et al., 2010). Downregulation of MidA in HEK294T cells results in a decreased level of Complex I. From Tocilescu's PhD thesis (2009), the Complex I content of R121K mutant in *Y. lipolytica* is only about 38% compared to WT, indicating that the assembly of Complex I is affected in this mutant.

Our collaborator Judy Hirst's group in Cambridge has proven that R121 in *Yarrowia* is also dimethylated (private correspondence), which is in accordance with what has been observed in another yeast species *P. pastoris* (Carroll et al., 2013), making it possible for us to do further study on this modification.

1.1.6 EPR spectroscopy theory

Electron paramagnetic resonance (EPR) or electron spin resonance (ESR) is widely used to detect the unpaired electron in paramagnetic species. In

1945, EPR phenomenon was discovered by the former Soviet Union physicist E. Zavoisky from chemicals which have paramagenetic properties, like $MnCl_2$, $CuCl_2$ etc. Since then, the development of EPR technique started to flourish and more advanced EPR approaches have been established. Because of its unique advantage compared to other spectroscopy, such as high sensativity, low sample volume etc., EPR has been widely applied to many fields involved in physics, cheimstry, biology, medical research etc.

EPR is a way to detect the energy transduction between different energy level when a paramagnetic substance is being excited in the provided magnetic field. Taking the free electron with a spin angular momentum S=±1/2 as an example, when applying an increasing magnetic field *H*, the spin energy level of the electron starts to split into two energy levels. This energy splitting is called electron Zeeman effect (Hagen, 2008; Brustolon, 2009) (Figure B8).

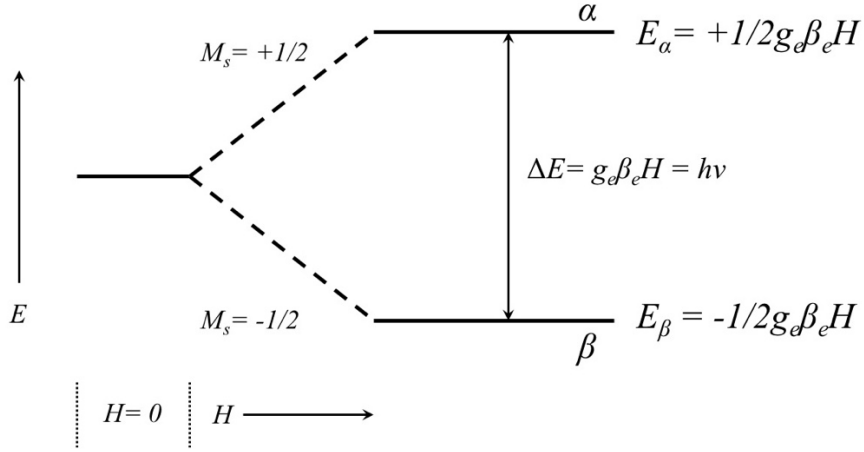


Figure B8 The energy splitting (Zeeman effect) of a free electron in an applied magnetic field *H*.

The higher level state *α*, with the magnetic quantum number M_s=+1/2, has the energy described as below:

$$E_\alpha = +1/2 g_e \beta_e H$$

Where:

β_e is the Bohr magnetón

g_e is the Landé factor (or simply called g factor) of a free electron

As an important EPR parameter, g factor reflects the contribution from electron spin angular momentum as well as the orbital angular momentum to its intrinsic magnetic moment. For a free electron which only carries the spin angular momentum, g_e=2.0023.

Similarly, the energy of the lower enegy level β with the magnetic quantum number M_s=-1/2 is:

$$E_\beta = -1/2 g_e \beta_e H$$

Thus, the difference between the two energy levels is:

$$\Delta E = E_\alpha - E_\beta = g_e \beta_e H$$

For a free electron, g_e and β_e are all constants. Therefore the energy separation ΔE is proportional to the applied external magnetic field H.

If an electromagnetic wave with the radiation frequency ν is perpendicularly applied to the H, and the electromagnetic radiation energy meets the below equation:

$$h\nu = g_e \beta_e H$$

The electron spin in the lower energy level β could then absorb this energy and transit to the higher energy level α. This resonance is thus called the electron magnetic resonance. The equation above is the foundation of EPR spectroscopy.

To achieve the resonance, in principle, we could either keep the magnetic field constant and change the frequency of the electromagnetic radiation ("frequency sweep"), or alternatively apply an electromagnetic wave with a constant frequency and change the magnetic field ("field sweep"). Due to the current limitation of the microwave techniques, the latter "field sweep"

method is widely applied to most of the EPR spectrometers. X-band is the most commonly used microwave band in EPR at a frequency of 9.75 GHz. In this case, The resonance field at g=2 is 0.348 T.

In this chapter, X-band continuous wave EPR (CW-EPR) is used to study cluster N2 from different Complex I mutants.

1.1.7 Aims of this research

Some studies have suggested that the undetectable N2 EPR signal in the *T. thermophilus* Complex I is probably caused by the low potential of cluster N2 (Hinchliffe et al., 2006; Meinhardt et al., 1990) and the adaption to use menaquinone instead of ubiquinone (Mooser et al., 2006). Another observation worth noticing is that ubiquinone is switched from menaquinone to semiquinone under anaerobic conditions, suggesting a possible explanation that the potential of cluster N2 in *E. coli* is lower than those in other species (Ingledew & Poole, 1984). Table B5 compares the reported N2 midpoint potentials from different species.

Table B4 The reported N2 midpoint potential value from different species

Species	Em (mV) at pH7	References
E. coli	-220	Leif et al., 1995
T. thermophilus	N2 not observed	Meinhardt et al., 1990; Hinchliffe et al., 2006
Y. lipolytica	WT: -140 H226M: -220	Zwicker et al., 2006; Le Breton et al., 2007
B. taurus	-140	Ingledew and Ohnishi, 1980

Investigating how Arg121 affects the potential of cluster N2 is the main topic of this chapter. Site-directed mutagenesis is employed to mutate Arg121 into either a similar residue or a different one. To study the dimethylation property of Arg121, the best way is to find the homologous gene of human *NDUFAF7* in *Y. lipolytica* and then erase the modification by knocking this gene out.

Combined with the above molecular biology techniques, EPR spectroscopy will be the ultimate approach to study cluster N2 in this project, providing more information on the local environment of cluster N2 and possibly expanding our knowledge of the function of cluster N2.

1.2 Materials and methods

1.2.1 Materials

1.2.1.1 Equipment

Incubator:

Excella E25 shaker (New Bruswick)

Innova 44R shaker (New Bruswick)

Centrifuge:

Centrifuge 5424 (eppendorf)

Centrifuge 5810R (eppendorf)

Universal 320R (Hettich)

Avanti J-30I equipped with JA1000 rotor (Beckman Coulter)

Optima™ L-100XP Ultracentrifuge equipped with Ti 45 or Ti 70 rotor (Beckman Coulter)

Column:

Self-packed Ni-NTA column (50 mL, filled with Ni Sepharose™ 6 Fast Flow, GE Healthcare)

Superose™ 6 Increase 10/300 GL (GE Healthcare)

Disrupter:

EmulsiFlex-C3 (Avestin)

DYNO®-MILL (Glenmills)

Spectrometer:

Cary 100 UV-Vis spectrophotometer (Aglient)

WPA Biowave II Spectrophotometer (Biochrom)

EPR spectrometer: Elexsys E500 spectrometer (Bruker) equipped with ER 4118X-MS2 resonator (Bruker);

Protein purifier: AKTApurifier™ (GE Healthcare)

pH meter: 3510 pH Meter (JENWAY)

Autoclave: tactrol 2 (Priorclave)

Electroporator: Gene Pulser Xcell™ (Biorad) equipped with Gene Pulser® Cuvette (Biorad, 0.2 cm gap)

Electrphoresis: Mini-PROTEIN® Tetra System (Biorad)

Thermocycler: T100™ Thermal Cycler (Biorad)

Peristaltic pump: P-1 (Pharmacia Fine Chemials)

Concentrator: Amicon® Utra-15 Centrifugal Filters (Merk Millipore)

1.2.1.2 Chemicals

Bovine Serum Albumin (BSA), Benzamidine hydrochloride hydrate, Tris (Trizma® Base), imidazole, Potassium ferricyanide (III), 6-aminocaproic acid, Lyticase, Menadione, PMSF, MOPS sodium salt, HEPES, Glycerol, Sodium phosphate monobasic, Lithium acetate dihydrate, NADH, Coenzyme Q1 (Sigma-Aldrich);

CHAPS (Alfa Aesar);

EDTA (VWR CHEMICALS);

Sodium dithionite (Merck KGaA);

Asolectin (Avanti® Polar Lipids);

n-Dodecyl β-maltoside (DDM) (GLYCON Bioch);

Decylubiquinone (UQ), Piericidin A, Rotenone (CAYMAN CHEMICAL);

cOmplete, EDTA-free protease inhibitor cocktail tablets (ROCHE);

Sodium chloride (Fisher Chemical);

Agar, YPD broth, LB broth, sorbitol (Formedium).

1.2.1.3 Antibiotics

Ampicillin sodium, Hygromycin B (Formedium)

1.2.1.4 Kits

E.Z.N.A® Yeast DNA Kit (OMEGA BIO-TEK, D3370);

E.Z.N.A.® Gel Extraction Kit (OMEGA BIO-TEK, D2500);

QIAprep Spin Miniprep Kit (QIAGEN, 27104);

Pierce® BCA Protein Assay Kit (Thermo Scientific, 23225);

In-Fusion® HD Cloning Kit (Takara).

1.2.1.5 Enzymes

OneTaq MasterMix, Q5 Polymerase MasterMix, T4 ligase (NEB);

Restriction Enzyme: NheI-HF, EcoRI-HF, BamHI, DpnI (NEB).

1.2.1.6 Plasmids

Name	Insertion	Source
pINA443	*YlURA3*	U. Brandt
pUC19-*YALI0D12859g*	*YALI0D12859g*	this work
pUC19-Δ*YALI0D12859g::YlURA3*	*YALI0D12859g* partially replaced by *YlURA3*	this work
pUB26-NUCM	*NUCM*	J. Hirst
pUB26-NUCM-R121M	*NUCM* carrying R121M mutation	J. Hirst
pUB26-NUCM-R121K	*NUCM* carrying R121K mutation	this work

1.2.1.7 Bacterial trains

E.coli: Stellar™ Competent Cells (Clontech);

Y. lipolytica: WT (GB10 background); *Δnucm* (GB10 background).

1.2.2 Methods

1.2.2.1 Molecular biology related methods

A. Site-directed mutagenesis (SDM)

The pUB26-NUCM (WT) and pUB26-NUCM-R121M plasmid as well as the corresponding *Y. lipolytica* strains were kindly given from Hirst's Group (MBU, Cambridge). The pUB26-NUCM-R121K mutant was made based on the above plasmid using the following site-directed mutagenesis method by introducing lysine (AAG) to replace the arginine 121 site in *NUCM* CDS.

1) The Primer design

The method to achieve the site-directed mutagenesis is generally replicating the whole plasmid by using a pair of primers near the expected

mutated site. In general, the primers were designed to have 3' "G/C" clamps and were modified with 5'-phosphorylation. In this case, the forward primer contains the mutated nucleotides (AAG instead of CGA). The reverse primer is a short sequence coming from the upstream of the mutated site. The primer sequences are listed in Appendix VI.

2) PCR

Q5 polymerase (NEB) was chosen to replicate the whole plasmid carrying the point mutation. The typical elongation time is ~40 s/kb. The reaction system and condition are shown as follows. If necessary, gradient annealing temperature can be attempted to optimize the best condition. When the PCR was finished, run a 1% TAE agarose gel at 80 V to make sure a visible PCR product can be generated (gel extraction is not necessary).

25 μL reaction system:

5×Q5 buffer (NEB)	5 μL
10 mM dNTPs	0.5 μL
10 μM Fwd Primer	1.25 μL
10 μM Rev Primer	1.25 μL
Plasmid (20 ng/μL)	1 μL
H_2O	15.75 μL
Q5 DNA polylmerase (NEB)	0.25 μL

PCR program:

1)	98 °C	00:30	
2)	98 °C	00:10	
3)	60-70 °C	00:30	
4)	72 °C	05:30	go to step 2 (24×cycles)
5)	72 °C	05:00	
6)	4 °C	Pause or forever	

3) Ligation

The below reaction was prepared in a PCR tube and incubated for 2 hours at RT (or using a thermal cycler at 25 °C). Then the sample was heated at 65 °C for 20 min to inactivate T4 ligase using a thermal cycler.

PCR product (directly from the PCR reaction)	5 μL
10×T4 DNA ligase buffer (NEB)	2 μL
H_2O	15 μL
T4 DNA ligase (NEB)	1 μL

4) DpnI digestion

1 μL DpnI was added to the ligation solution at 37°C for 15 min. After digestion, the reaction was heated at 80°C for 20 min to inactivate DpnI using a thermal cycler.

5) Transformation

All the above solution was transferred to the Stellar™ Competent Cells following manufacturer's instruction (Clontech, protocol PT5055-2). The transformed cells were spread onto ampicillin-containing plates at 37 °C overnight.

6) Sequencing

Several colonies were picked from the above plates to make small culture and then the plasmid was extracted for sequencing. Primers used for sequencing the pUB26-NUCM plasmid (including the one carrying the point mutation) are shown below:

Sequencing the NUCM insertion direction: n49knc1; 49-2.5;

Sequencing the CDS: NUCM-R (reverse); 49-1 (forward).

The detailed information of the primer sequence is in Appendix VI.

B. Gene cloning

The gene cloning was conducted in a similar way to what's described in part A of 1.2.2.1 to amplify the gene by Q5 polymerase, but extend the cycle number to 30.

The gene insertion into pUC19 plasmid and the replacement of the target gene by *YlURA3* gene were all achieved by the In-Fusion cloning technology using In-Fusion® HD Cloning Kit (Takara).

In brief, the gene-specific primer for gene amplification is designed with an extra 15 bp extension that is homologous to the end of the linearized vector. By PCR, each end of the gene was added with this 15 bp vector sequence. The In-Fusion enzyme could fuse this gene fragment and the linearized vector by homologous recombination from the overlaps, thus achieving the gene insertion into the vector.

The detailed information of the primer sequence is in Appendix VI.

C. Yeast genome extraction

The genome extraction from yeast cells was performed following manufacturer's instruction of the E.Z.N.A® Yeast DNA Kit (OMEGA BIO-TEK).

D. Yeast plasmid extraction

Since yeast cells are hard to break, an extra cell wall digestion was needed. The cell pellet was collected from 3 mL yeast culture by centrifuge, and then resuspended in 500 µL TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) containing 40 µL 5,000 unit/mL lyticase (SIGMA, L4025) at 30°C for 1h to digest the cell wall. After lyticase digestion, the plasmid extraction was performed using the QIAprep Spin Miniprep Kit (QIAGEN) following manufacturer's instruction.

Note that the plasmid extracted by this method is not visible by DNA gel. But if transferred into *E.coli* cells, the plasmid can be further simply extracted at a notably amount and is qualified enough for either DNA gel or sequencing.

E. Yeast transformation and colony PCR

The transformation of the plasmid into the corresponding *Y. lipolytica* strain was followed by the method below slightly modified from Wang et al., 2011. Here is a brief procedure.

Yeast strain was inoculated into 20 mL starter cultures of 2×YPD and left shaking at 30°C overnight. When the OD_{600} of this starter culture reached to 13-18, another 100 mL of 2×YPD was inoculated using this 20 mL starter culture to an OD_{600} of 0.1 and incubated by shaking at 30°C overnight. When the OD_{600} of the 100 mL culture reached to 1-2.5, the cell pellet was harvested at 5000 g for 5 mins and resuspended into 8 mL competence buffer (10 mM Tris-HCl, pH 7.5, 0.6 M sorbitol, 150 mM Li-Acetate). The cells were incubated in competence buffer for 1 hour at R.T. with occasional mixing and then harvested again as before. Note that from now on, all subsequent steps should be done on ice or in the cold room. The cells were resuspended and washed in 3 mL ice-cold 1 M sorbitol. The washing step was repeated 3 times with the final resuspension into 5 mL ice-cold 1 M sorbitol. The cells were kept on ice untill ready for transformation.

In a cooled electroporation cuvette with a 2 mm gap, at least 1 μg of DNA (5 ug would be best) was mixed with 200 uL competent yeast cells thoroughly with the pipette. The mixture was electroporated with 1.5 kV, 200 ohms and 25 μF (Using the exponential protocol in the BIORAD electroporator. This protocol should give a pulse length at around 4.6 ms). Then 1 mL RT 2×YPD was immediately added to the cells. The cells were then transferred into a 1.5 mL Eppendorf tube and incubated for 2-3 hours by shaking at 30°C. After incubation, the cells were plated onto the required selection plates.

For the transformation of pUB26-derived plasmids, a final concentration of 150 μg/ml hygromycin was required in the selection medium. To make sure the plasmid was successfully transferred into the yeast cells, a control plate pasted with electro-shocked competent cells without plasmid was needed. Several relatively larger single colonies were picked respectively from the plate for colony PCR (Small colonies that were similar to what was growing in the control plate usually come from the background, but not positive transformants). Each single colony was resuspended in 40 μl TE buffer and then treated at 95°C for 20min in a PCR cycler. Then normal colony PCR was conducted using this crude sample as the template. If necessary, the mutated site can be again sequenced to make sure the yeast carries the correct point mutation. In this case, the plasmid is needed to be extracted from the yeast following the method described in section D from 1.2.2.1.

1.2.2.2 Microbiology and protein related methods

A. Standard large-scale *Y.lipolytica* growth procedure

A tiny amount of yeast glycerol stock was inoculated into 50 ml 2×YPD (pH 5.5) in a 250 ml conical flask. Hygromycin is added at a final concentration of 50 μg/mL in the medium if *Y.lipolytica* mutation is needed to grow. The flask was incubated at 30°C by shaking at 225 rpm until OD_{600} was around 10-20. 2 mL of this culture was transferred into a 2 L flask containing 500 mL 2×YPD, and left for shaking until the cell density reaches to 20-30. The cells were harvested at 5000 g for 15 min and kept at -80°C.

B. Protein quantification

The quantification of either isolated Complex I or total protein from the membrane is followed by the manufacturer's instruction of Pierce® BCA Protein Assay Kit (Thermo Scientific).

C. Membrane Preparation from *Y. lipolytica*

Required solutions:

Breaking buffer (pH 7.2): 20mM MOPS-Na, 400 mM Sorbitol, 5 mM EDTA, 5 mM Benzamidine hydrochloride hydrate, 5 mM 6-aminocaproic acid, 2 % BSA

Washing buffer (pH 7.2): 20mM MOPS-Na, 400 mM Sorbitol, 5 mM Benzamidine hydrochloride hydrate, 5 mM 6-aminocaproic acid

Resuspension buffer (pH 7.45): 20 mM NaH_2PO_4, 50 mM NaCl

1) Cell disruption

The yeast cell pellet was defrosted at 4°C overnight. The disruptor (Emulsiflex Avestin C3) was pre-chilled to 4°C. Approximately 50 g cells were resuspended in 100 mL ice-cold breaking buffer (containing 2 mL fresh 200 mM PMSF). The resuspended cells were disrupted at 25-30 Kps for ~15 rounds. The cells were frequently checked under microscopy.

Alternatively, if Dyno-mill is chosen to disrupt the cells, normally 500 g cells were resuspended in 1 L ice-cold breaking buffer (containing 4 mL fresh 200 mM PMSF) for the disruption. The cells were disrupted by pumping through the system twice at 8-10 rpm/min.

2) Ultracentrifugation

The disrupted cells were collected at 5700 g in 4°C for 30mins to get rid of the cell debris. The supernatant was transferred to the ultracentrifuge (Beckman Coulter) at a speed of 208,000 g for 1 hour at 4°C. The membrane pellet was collected and resuspended in ice-cold washing buffer using a glass homogenizer and then re-spin at 208,000 g at 4°C for 1 hour. This step was repeated to wash the membrane thoroughly. The pellet was finally resuspended in ice-cold resuspension buffer and divided into 20-25 mL aliquots then kept at -80°C.

D. General protein purification of Complex I from *Y. lipolytica* membrane

Required solutions:

2% asolectin/CHAPS stock (25mL): Weight out 500 mg CHAPS, and dissolve in 25 mL of buffer (containing 20mM NaH_2PO_4, 400 mM NaCl, 52 mM imidazole). This solution is homogenized using a glass homogenizer to make the 2 % stock (syringe through a 0.22 μm filter before use)

Buffer A (pH 7.2): 20mM NaH_2PO_4, 400 mM NaCl, 52 mM imidazole, 0.1% DDM, 0.02% asolectin/CHAPS (filtered and degassed before use; the filtered DDM and asolectin/CHAPS should be added freshly before loading to the column)

Buffer B (pH 7.2): 20mM NaH_2PO_4, 400 mM NaCl, 140 mM imidazole, 0.1% DDM, 0.02% asolectin/CHAPS (filtered and degassed before use; the filtered DDM and asolectin/CHAPS should be added freshly before loading to the column)

Gel Filtration buffer (pH 7.45): 20 mM Na-MOPS, 150 mM NaCl, 0.05% DDM, 10% glycerol (filtered and degassed before use; the filtered DDM should be added freshly before loading to the column)

Neutralised (pH 7.5) 1M imidazole

All the steps below should be performed on ice or in a cold room.

1) Membrane dissolution

Approximately 20-50 mL mitochondrial membrane pellet was thawed in RT water if frozen. The membrane was ultra-centrifuged at 208,000 g for 30 mins at 4°C. The pellet was homogenized in resuspension buffer (from part C of 1.2.2.2) to a finally concentration of 20 mg/mL. Two EDTA-free protease inhibitor cocktail tablets (ROCHE) were added into this solution. 10% DDM was added dropwise to a 2:1 ratio of DDM: protein by stirring. The membrane was solubilized for 30 mins. The supernatant was collected at 208,000 g at

4°C for 1 hour. The solid NaCl was added to a final concentration of 400 mM and imidazole was added to 20 mM from 1M neutralised stock.

2) Protein purification

The above sample was loaded onto a pre-equilibrated 50 mL Ni-NTA column which was connected to the AKTApurifier™. The impurities were washed by 100% Buffer A. Complex I was eluted by 100% Buffer B. All the fractions from the Complex I peak were collected and concentrated by using a Millipore 15 (50,000 MWCO) concentrator. The concentrated protein was injected onto the pre-equilibrated Superose 6 increase column at a flow rate of 0.5 mL/min. Complex I peak was then concentrated as described above to 5-20 mg/mL and frozen as 10-20 μL aliquots in dry ice or liquid nitrogen and kept at -80°C.

E. Potentiometric titration

Required solutions:

TIP 7 buffer: 50 mM HEPES, 50 mM Na_2HPO_4, 150 mM NaCl, 0.03% DDM, 30% glycerol

5/10/20/40 mM gradient sodium dithionite

5/10/20/40 mM gradient $K_3Fe(CN)_6$ solutions

The potentiometric titration of Complex I by the reductant sodium dithionite (SDT) or the oxidant $K_3Fe(CN)_6$ under anaerobic condition is mainly followed what's described in Wright et al., 2016 (Figure B9). Here is a brief description.

First, the protein was exchanged into TIP7 buffer using the Millipore 15 (30,000 MWCO) concentrator at 4°C. Then the below apparatus was prepared in the glovebox where an anaerobic environment was maintained. The apparatus was chilled to 4°C by connecting to a water bath. The electrodes connected to the PSTrace 4.6 Software were assembled onto the cell in order to record the current potential. All the redox potentials of the

sample recorded here were relative to the Ag/AgCl reference electrode and were needed to convert to the standard hydrogen electrode (SHE). At least 140 μl protein sample was added into the cell with two tiny stirring balls (2mm diameter). The following mediators were added to the protein as well: Indigotrisulfonate, Methylene Blue, Benzyl Viologen, Indigodisulfonate, Anthraquinone-2-sulfonic acid. The final concentration of each mediator was the same as the protein. The solution was stirred at full speed for at least 15 min for degassing.

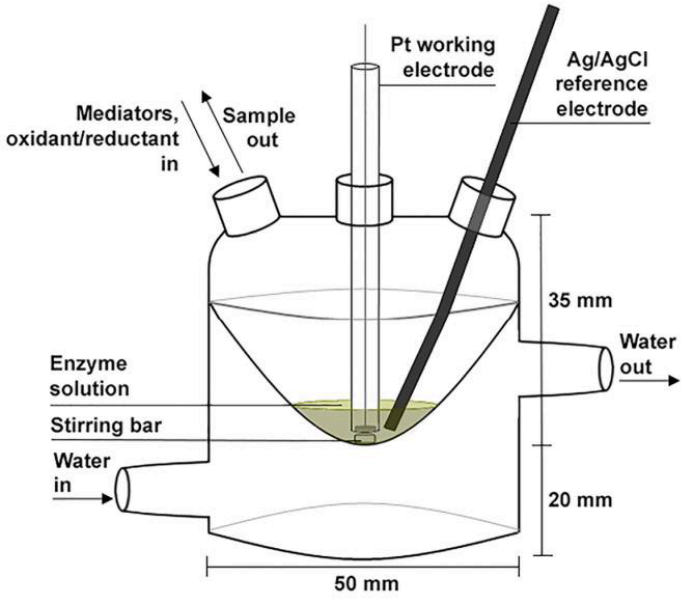


Figure B9 The schematic of the apparatus for the potentiometric titration (© 2016 Wright et al., by permission)**.**

Several SDT and $K_3Fe(CN)_6$ solutions with gradient concentration (5-40 mM) were made respectively in the glovebox. Substoichiometric amounts of either the reductant or oxidant were added to the protein samples to achieve different degrees of potential. Each sample (~10 μL) with a certain potential was transferred into the EPR tube and fast-frozen in the liquid nitrogen. The samples were ready to be measured by EPR.

G. EPR spectroscopy

All the EPR samples were measured in X-band continuous-wave mode (CW) using an ER 4118X-MS2 resonator (Bruker) in an Elexsys E500 spectrometer (Bruker) equipped with a cryostat system (Cryogenic Ltd., UK). The detailed EPR condition for each measurement is shown in the legend of the corresponding spectrum in the result part.

For the titrated samples, the sample without adding any reductant/oxidant was used as the background. For all the other individual samples, the corresponding buffer was used as the background.

The EPR data were analyzed and simulated by EasySpin software package by Matlab (Stoll & Schweiger, 2006).

1.3 Results

1.3.1 The measurement of N2 redox potential in wide type Complex I from *Y. lipolytica*

The purification of Complex I from *Y. lipolytica* has been established in our lab. The His-tagged wild type Complex I was successfully purified by Ni-affinity chromatography followed with gel filtration.

In order to make sure that the purified Complex I functions well, the N2 redox potential of *Y. lipolytica* wild type Complex I was measured. The reported N2 midpoint potential ($E_{m,7}$) of the wild type Complex I in *Y. lipolytica* at pH 7 is -140 mV (vs Standard Hydrogen Electrode (SHE)) (Zwicker et al., 2006). Other reported value from various species can be found in Table B5 in section 1.1.7.

The potentiometric titration of the wild type protein was carried out under anaerobic conditions in the glovebox. The protein potential was controlled by adding substoichiometric amounts of sodium dithionate (SDT) and the plot in Figure B10a below shows the recorded potential data (vs SHE). The samples with different degrees of potential were measured by CW-EPR (Figure B10b). Then the EPR data was simulated and the integral area of $N2_z$ was used to fit to the one-electron Nernst curve (Figure B10c). The result shows that the $E_{m,7}$ of N2 from the wild type Complex I is -137.8±4 mV, which is in line with the reported data (Zwicker et al., 2006), indicating that the purification method used here works well.

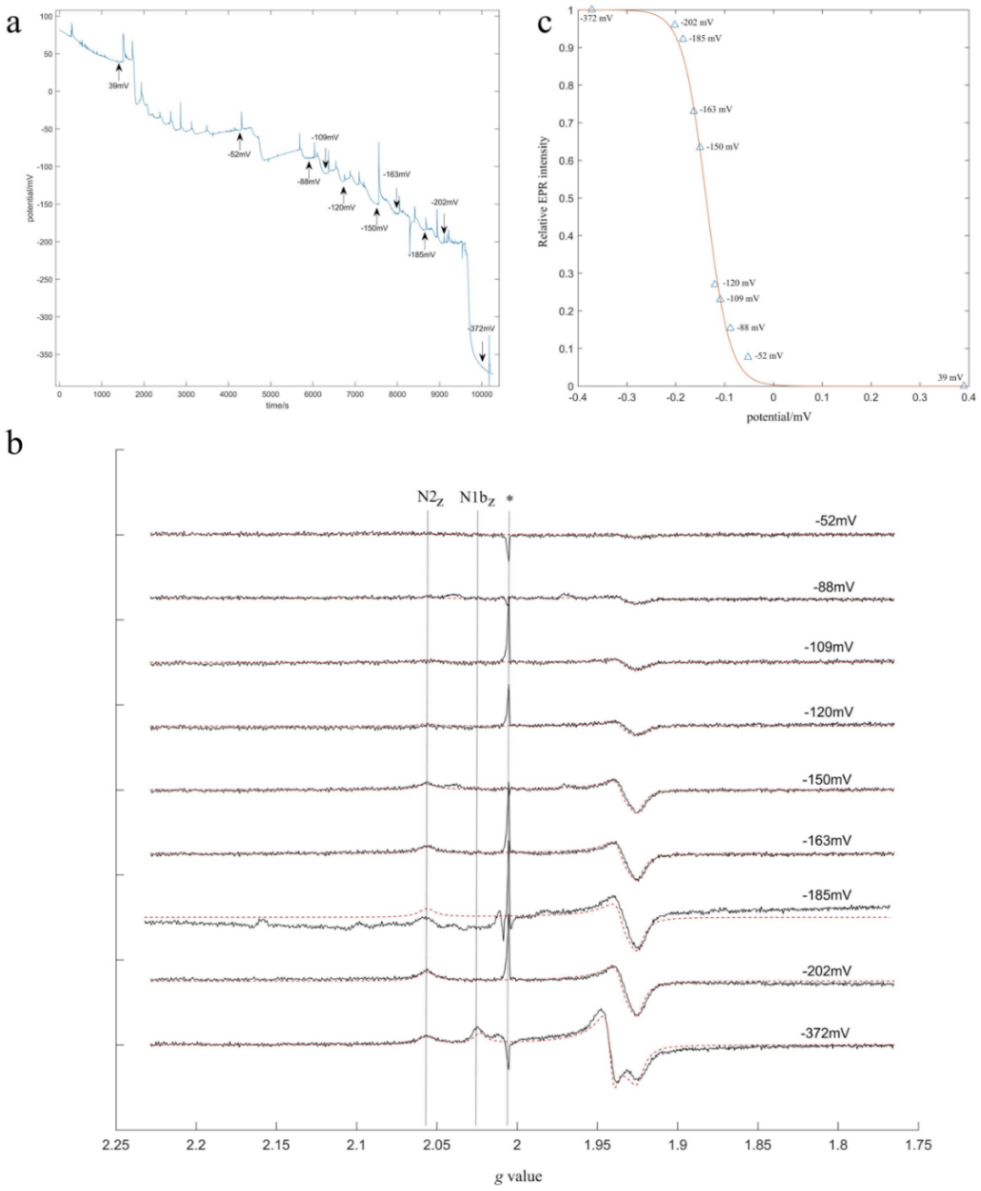


Figure B10 The measurement of N2 redox potential in the wild type Complex I from *Y. lipolytica*. a. The titration trace of the WT Complex I potential (vs SHE) by adding different amount of sodium dithionate at 4°C under anaerobic conditions. b. The CW-EPR spectra of the samples from the titrations. The black spectra indicate the experimental data. The red dashed spectra indicate the simulated spectra. The vertical dotted lines indicate the corresponding Fe-S cluster. The asterisk indicates the signal of unknown impurity from the EPR tube. EPR measurement condition: temperature, 25 K; microwave frequency, 7 GHz; microwave power, 2 mW. c. The one-electron Nernst curve of the titration data. The integral area of $N2_z$ from Figure b is used to fit for the curve.

1.3.2 The research of the R121M variant

The R121M mutation in *NUCM* gene from *Y. lipolytica* was first studied by EPR. The R121M variant was successfully purified by Ni-affinity chromatography followed by gel filtration. The SDS-PAGE gel indicates the clear bands of each core subunit compared with WT protein (Figure B11).

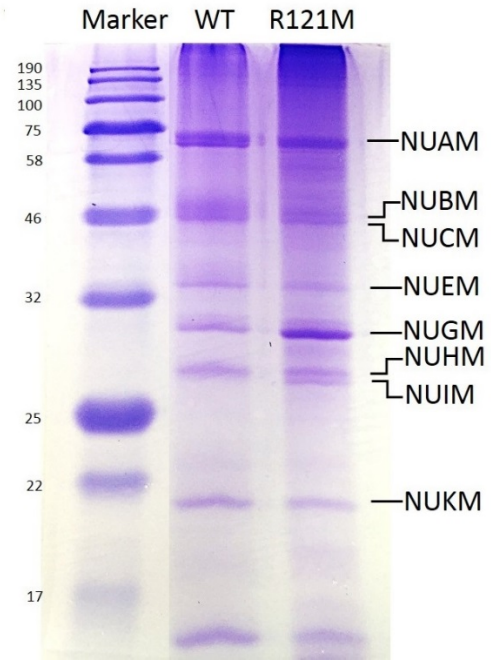


Figure B11 SDS-PAGE gel of WT and R121M Complex I from *Y. lipolytica*.

The SDT-reduced R121M protein was then measured by CW-EPR. Compared to the WT spectrum, no N2 signal is visible in the R121M protein (Figure B12 bottom spectrum). The isolated spectrum of individual Fe-S cluster has also been simulated. Again, we cannot simulate any N2 signal contributing to the overall R121M spectrum, indicating that the absence of this residue results in a severe effect on cluster N2. The lack of N2 signal also makes it impossible to measure the midpoint potential of N2 in this mutant. The *g* value of each cluster has been summarized in Table B6 together with previously published data. The *g* values from this work are consistent with the literature.

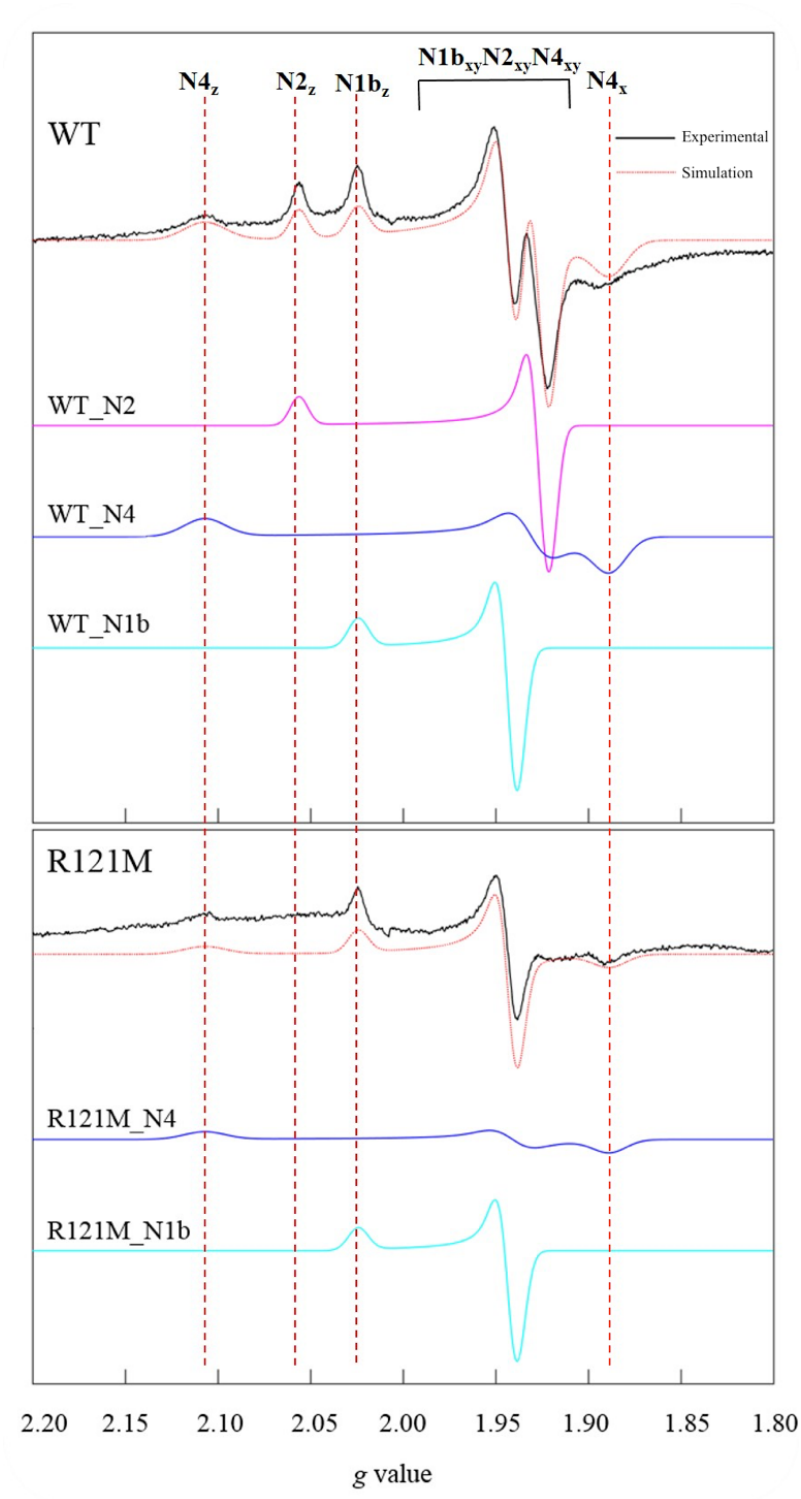


Figure B12 CW-EPR spectra of fully reduced WT and R121M Complex I. Experimental spectra of fully reduced R121M Complex I protein are shown in black. Simulated spectra are shown in red. All samples were reduced by SDT. EPR measurement condition: temperature, 15 K; microwave frequency, 7 GHz; microwave power, 2 mW.

Table B5 *g* values from the simulated CW-EPR spectra in this work and the comparison with published data

Fe-S cluster	*g* values			*g* values			*g* values			*g* values		
	g_z	g_y	g_x	g_z	g_y	g_x	g_z	g_y	g_x	g_z	g_y	g_x
N1	2.020	1.943	1.94	2.018	1.945	1.933	2.02	1.95	1.93	2.024	1.943	1.939
N2	2.053	1.929	1.925	2.051	1.926	1.918				2.056	1.928	1.920
N3	2.030	1.924	1.860	2.031	1.930	1.861						
N4	2.103	1.939	1.894	2.104	1.931	1.892	2.10-2.11	1.93	1.89	2.107	1.931	1.888
N5	2.065	(1.93)	(1.90)	2.062	1.93	~1.89	2.06	1.93	1.89			
Reference	Djafarzadeh et al., 2000			Kerscher et al., 2001; Kerscher et al., 2002; Garofano et al., 2003			Summarized by Yakovlev et al., 2007			This work		

From the observation, two hypotheses were proposed to explain the loss of the N2 signal in R121M. The potential of R121M could probably decrease to a much lower level that SDT could not achieve. Then we tried to use EuDTPA as a more powerful reductant to lower the protein potential as much as possible (Reda et al., 2008; Roessler et al., 2012), but failed. Another idea is that the spin state of N2 is possibly changed to a higher degree in the R121M variant. To test if there are any high-spin signals of N2 in R121M, the mutated protein was measured under a much lower magnetic field. Unfortunately, nothing with high-spin signals was found by background subtraction (Figure B13).

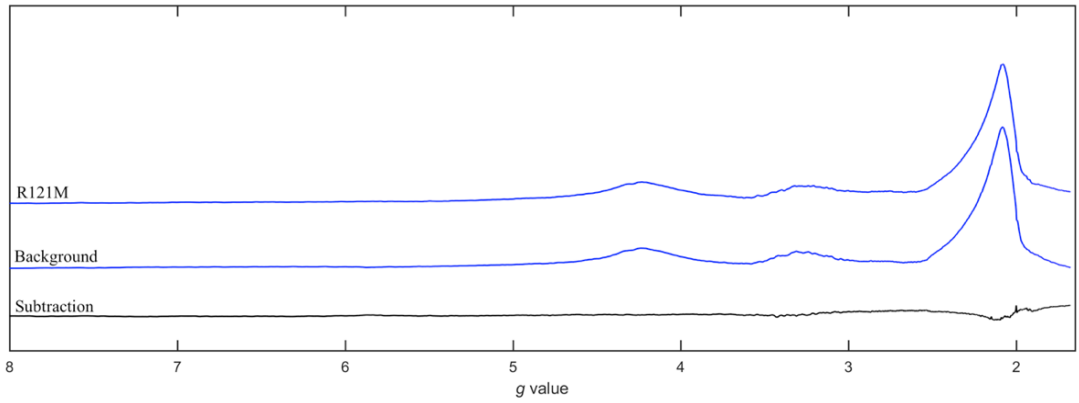


Figure B13 Experimental CW-EPR spectra of fully reduced R121M Complex I at lower magnetic field. EPR measurement condition: temperature, 5 K; microwave frequency, 7 GHz; microwave power, 20 mW.

1.3.3 The research on the R121K variant

Due to the severe effect from the R121M mutant, another variant R121K was chosen for our next study. The structure of lysine is most close to arginine, and both side chain groups are positively charged. Thus, the R121K mutation was expected to have a mild effect on cluster N2.

The R121K mutation in *Y. lipolytica* has been previously reported. The Complex I content in mitochondrial membrane carrying the R121K mutation is only 38±2% compared to the WT. The EPR characterizations of this mutant have only been reported from the crude samples (the entire mitochondrial membrane containing the mutated complex) and no N2 signal was detected (Tocilescu, PhD thesis, 2009). To have a clearer idea, this mutant was made in this study again and the isolated protein was purified for EPR investigation.

The mutant has been successfully generated by the method described in section 1.2.2.1. The complex was then expressed and purified. Unfortunately, the CW-EPR spectrum of the isolated R121K Complex I again shows no N2 signal (Figure B14), similar to that of the R121M variant.

The study on the R121M and R121K mutants both indicates the importance of R121 residue for N2 function. Apparently R121 cannot be simply replaced by a similar residue to complement the deficiency. Due to lack of the EPR signals from both mutants, advanced pulse EPR approach cannot be employed for our further investigation. Therefore, the study of the dimethylation of R121 residue became the next interest of this project.

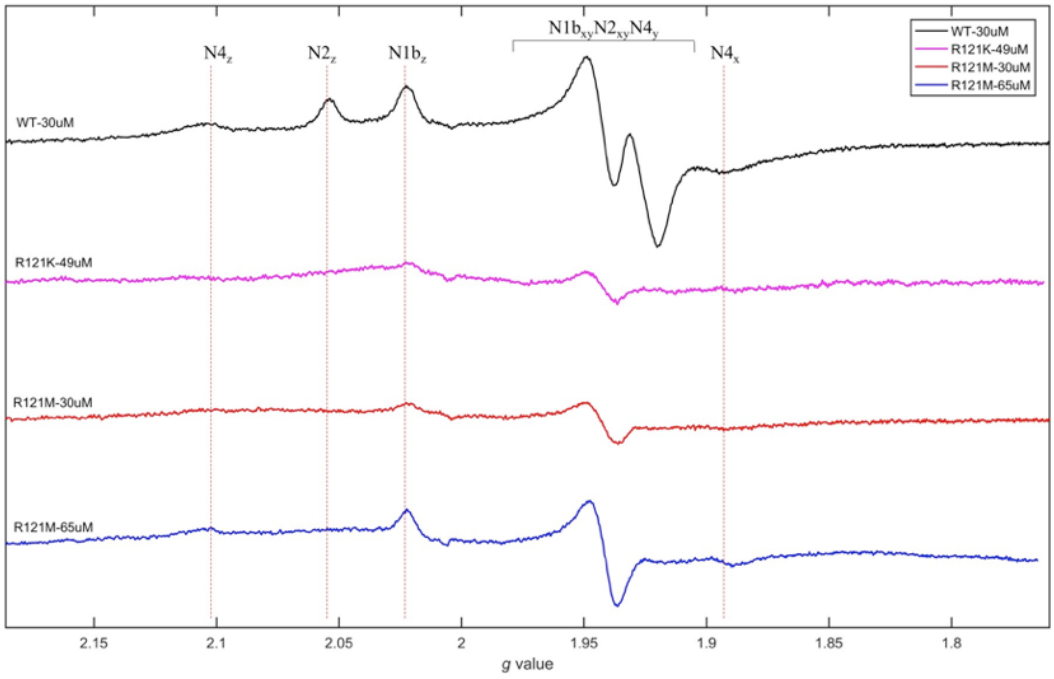


Figure B14 Experimental CW-EPR spectra of fully reduced R121K Complex I compared with WT and R121M at 15K. EPR measurement condition: microwave frequency, 7 GHz; microwave power, 2 mW.

1.3.4 Making the NDUFAF7 homolog gene knock-out in *Y. lipolytica*

1.3.4.1 *YALI0D12859g* is homologous to human NDUFAF7 in *Y. lipolytica* by bioinformatic analysis

As discussed in the introduction, human NDUFAF7 is the methyltransferase responsible for dimethylating residue R85 in the human NDUFS2 subunit (corresponding to NUCM subunit in *Y. lipolytica*) (Rhein et

al., 2013). To find the homologs of human NDUFAF7 in *Y. lipolytica,* several bioinformatics analyses were attempted by sequence alignment from different databases, briefly described below:

1) Using NDUFAF7 protein sequence as query

When using NDUFAF7 as the keyword, two *Y. lipolytica* proteins were found from the Uniprot database, Q6C0X9 and Q6C998. If using human NDUFAF7 protein sequence as query in the KEGG *Y. lipolytica* genome database(*http://www.genome.jp/kegg-bin/show_genomemap_top?org_id=yli*), the most likely hit is yli:YALI0D12859g given that it has the highest score. This sequence is annotated as Q6C998.

2) Using *NDUFAF7* nucleotide sequence as query

To further search for the sequences within larger database in NCBI, the human *NDUFAF7* (NCBI Reference Sequence: NM_144736.4) gene CDS sequence was used as query in the *Y. lipolytica* (taxid:4952) database by BLASTX. Only two protein sequences are found, YALI0D12859p and YALIA101S12e01288g1_1. They are only differed by two residues. When aligning the sequences in pair, YALI0D12859p and Q6C998 are exactly the same.

If using the same human *NDUFAF7* gene sequence as query in the *Y. lipolytica* CLIB122 (taxid:284591) database, only YALI0D12859p was found and it makes no difference with previous results shown above. Using the whole mRNA sequence of *NDUFAF7* gave the same results. In conclusion, from the nucleotide sequence of human *NDUFAF7* gene, only one protein Q6C998 was found.

From the above information, Q6C998 encoded by *YALI0D12859g* is most likely the homologous protein of human NDUFAF7 in *Y. lipolytica*. Sequence alignment shows that the two proteins share 31.9% identity and 47.6% similarity (by EMBOSS Needle, Matrix: EBLOSUM62) (Figure B15).

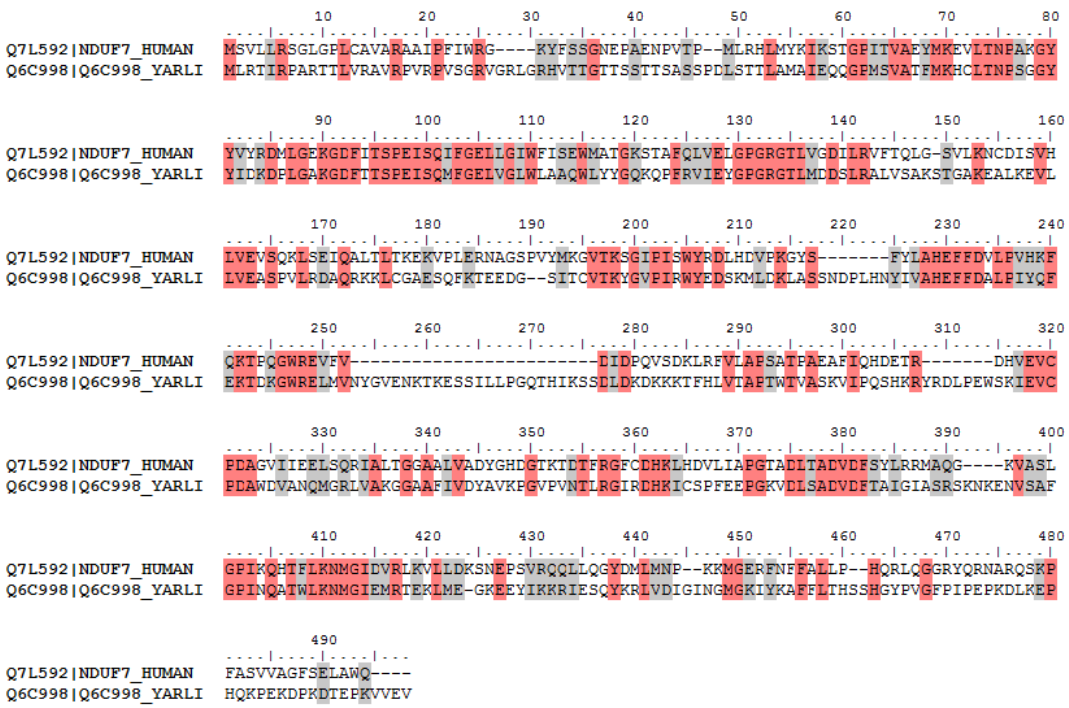


Figure B15 Sequence alignment of NDUFAF7 and Q6C998 protein. The identical residue pairs from both proteins are indicated in red shade while the residue pairs sharing high similarity (>80%) are indicated in grey shade.

1.3.4.2 The method to make the *NDUFAF7* homolog gene knock-out in *Y. lipolytica*

To knock out the NDUFAF7 homolog gene in *Y. lipolytica,* the *URA3* gene encoding Orotidine 5'-phosphate decarboxylase (ODCase) was chosen to replace the target gene *YALI0D12859g*. *URA3* gene has been widely used in yeast engineering. ODCase plays an important role in the synthesis of pyrimidine ribonucleotides (Flynn & Reece, 1999). Loss of ODCase activity severely affects the normal growth unless uracil or uridine is added to the media, making it easier for the screening by this auxotrophy. Figure B16 briefly describes the method designed for replacing the target gene with the marker gene by homologous recombination.

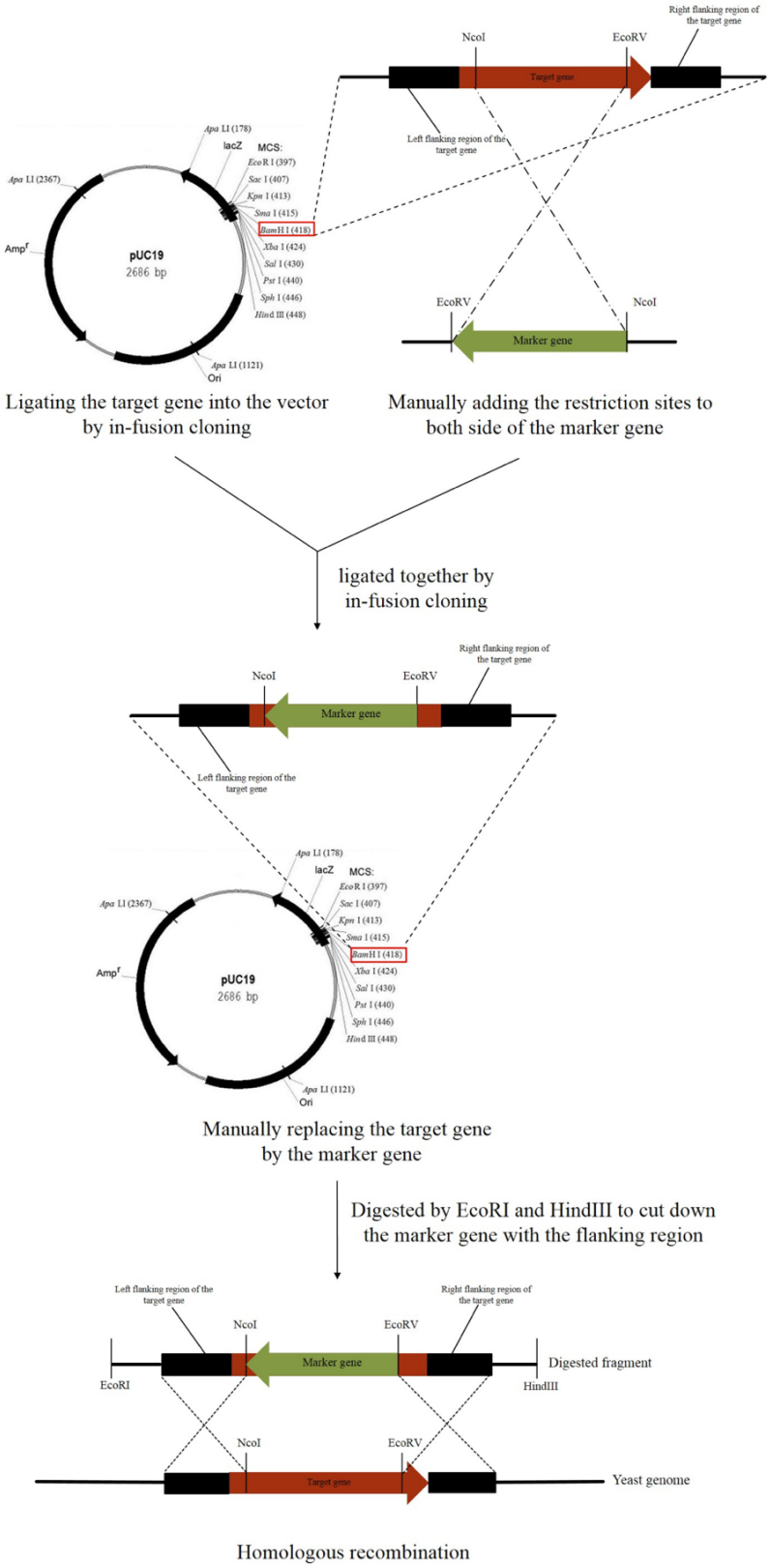


Figure B16 Schematic representation of the method to replace the target gene with marker gene by homologous recombination. In this study, the marker gene is *URA3*, and the target gene is *YALI0D12859g*.

Following the original method reported from Brandt's group (Kashani-Poor et al., 2001) (using *URA3* to replace *NUGM*), the target gene *YALI0D12859g* together with its ~1000bp upstream sequence and ~1000bp downstream sequence was cloned and then inserted into the pUC19 plasmid at the BamHI site. Since NcoI and EcoRV are located near the beginning and the end of the ORF of the target gene respectively, they were chosen to double digest this gene in order to cut off most of the ORF (Figure B17a, the 1.2Kb fragment). Then the marker gene *YlURA3* (originated from plasmid pINA443) was reversely inserted into this gap.

The whole map of this newly generated plasmid is shown in Figure B17c. The 4kb Δ*YALI0D12859g::YlURA3* fragment was cut off from the plasmid by EcoRI and HindIII (Figure B17b) from the plasmid. This 4K fragment is used for the transformation into *Y. lipolytica* to achieve the homologous recombination. All the ligations were done by in-fusion cloning.

Though the plasmid has been successfully generated, unfortunately, the transformation into *Y. lipolytica* gave no positive strains by several trials. Even for those strains with *URA3* gene insertion, the marker gene is off-targeted to other gene loci rather than *YALI0D12859g*. Apart from the off-target problem, another explanation is that *YALI0D12859g* may have very important function to the cell survival, thus knocking out the gene would be lethal to *Y. lipolytica*.

In all, the correct gene knockout strains were not obtained from this study.

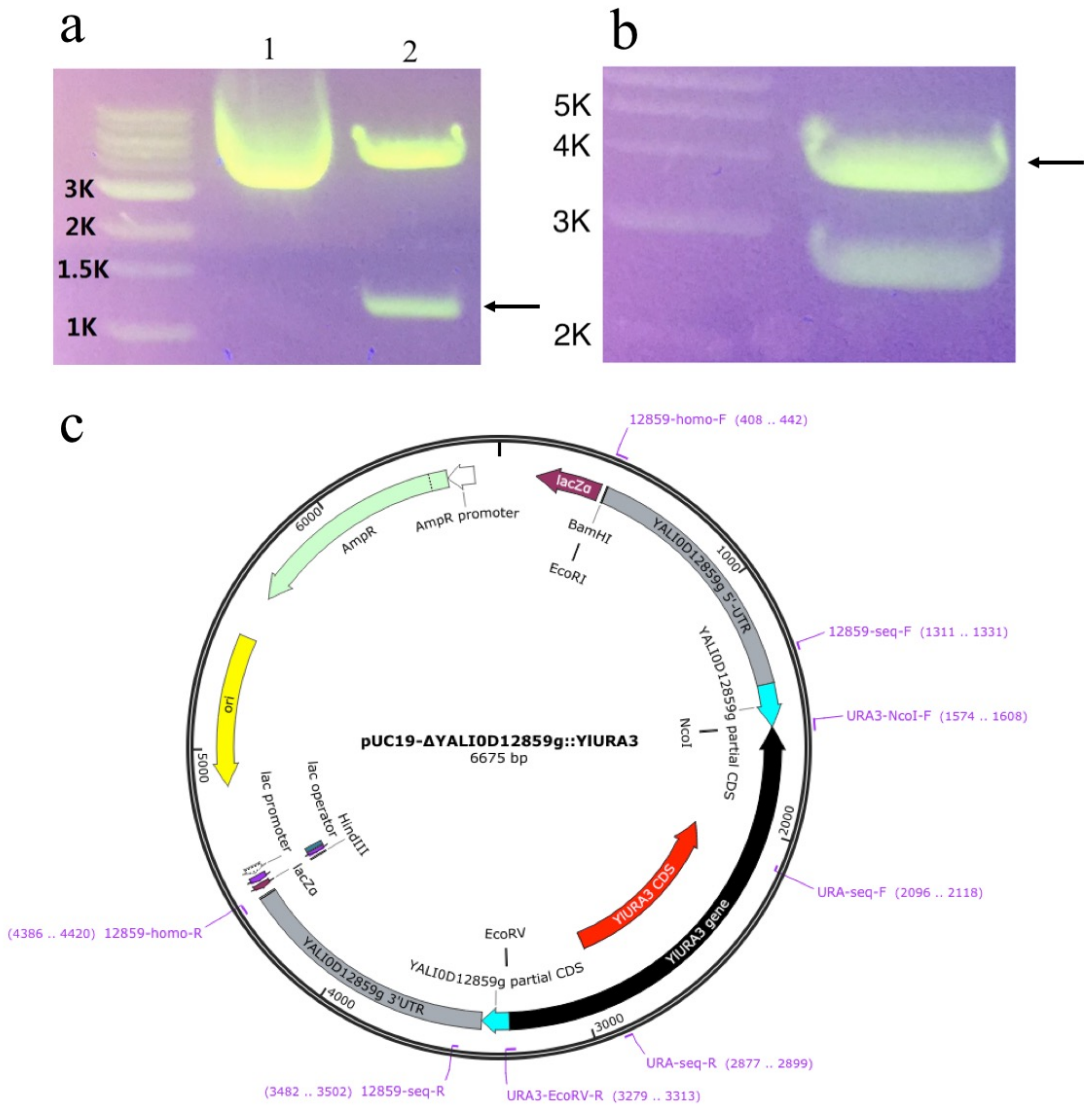


Figure B17 The making of pUC19-*ΔYALI0D12859g::YlURA3* plasmid. a. Double digestion of pUC19-*YALI0D12859g*. Lane 1, control plasmid without treatment. Lane 2, plasmid digested by NcoI and EcoRV. Arrow indicates the 1.2Kb ORF fragment. b. Double digestion of pUC19-Δ*YALI0D12859g::YlURA3* plasmid by EcoRI and HindIII. Arrow indicates the 4Kb Δ*YALI0D12859g::YlURA3* fragment. c. Plasmid map of pUC19-Δ*YALI0D12859g::YlURA3*.

1.4 Conclusion and discussion

In this chapter, the function of R121 residue in NUCM subunit from Complex I is studied by combining site-directed mutagenesis with EPR spectroscopy.

By CW-EPR spectroscopy, the signal of cluster N2 is absent from both R121M and R121K mutations and the possibility that cluster N2 changes to a higher spin state has been excluded. These findings indicate the importance of R121 to maintain the normal function of N2 in Complex I. Unfortunately, the lack of EPR signal of N2 cluster from both R121M and R121K variants makes it impossible for further advanced EPR studies.

We then turned to focus on the possible effect of dimethylation state of residue R121 on N2 cluster. To erase the dimethylation on this residue, the gene responsible for this modification was searched by bioinformatics. Homologous recombination method is chosen to knock this gene out. Most of the CDS sequence of this gene (*YALI0D12859g*) has been replaced by the marker gene *YlURA3* and inserted into the pUC19 plasmid. Unfortunately, by several transformations, the gene knockout is still unsuccessful with a high off-target rate observed. Since previous reports have demonstrated that this gene has great effects on complex I assembly or stability (Carilla-Latorre et al., 2010; Rhein et al., 2013), the deficiency of this gene would probably greatly affect the yeast viability.

Instead of the conventional homologous recombination method to knock this gene out, deleting part of the gene or making frameshift mutation by the CRISPR-CAS9 technique would probably be an alternative way.

To conclude, R121 from NUCM subunit is one of the key residues that affect the direct environments around cluster N2. However, appropriate methodology is required to further investigate its function in mitochondrial Complex I.

The above work was done by the early of 2017. Until recently, the EPR data from this work was published as part of the support to the cryo-EM structure of NUCM-R121M complex I (Hameedi et al., 2021). In this structure, the R121M mutation causes the missing of the N2 cluster. The nearby quinone-binding site is also disordered. It indicates the key roles of R121 in stabilizing the N2 cluster.

1.5 Reference

Baradaran, R., Berrisford, J. M., Minhas, G. S., & Sazanov, L. A. (2013). Crystal structure of the entire respiratory complex I. *NATURE*, 494(7438), 443-448.

Bedford, M. T. (2007). Arginine methylation at a glance. *JOURNAL OF CELL SCIENCE*, 120(Pt 24), 4243-4246.

Bedford, M. T., & Clarke, S. G. (2009). Protein arginine methylation in mammals: who, what, and why. *MOLECULAR CELL*, 33(1), 1-13.

Berrisford, J. M., & Sazanov, L. A. (2009). Structural Basis for the Mechanism of Respiratory Complex I. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 284(43), 29773-29783.

Boffa, L. C., Karn, J., Vidali, G., & Allfrey, V. G. (1977). Distribution of NG, NG,-dimethylarginine in nuclear protein fractions. *Biochem Biophys Res Commun*, 74(3), 969-976.

Carilla-Latorre, S., Gallardo, M. E., Annesley, S. J., Calvo-Garrido, J., Grana, O., Accari, S. L., Smith, P. K., Valencia, A., Garesse, R., Fisher, P. R., & Escalante, R. (2010). MidA is a putative methyltransferase that is required for mitochondrial complex I function. *JOURNAL OF CELL SCIENCE*, 123(Pt 10), 1674-1683.

Carroll, J., Ding, S., Fearnley, I. M., & Walker, J. E. (2013). Post-translational modifications near the quinone binding site of mammalian complex I. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 288(34), 24799-24808.

Djafarzadeh, R., Kerscher, S., Zwicker, K., Radermacher, M., Lindahl, M., Schägger, H., & Brandt, U. (2000). Biophysical and structural characterization of proton-translocating NADH-dehydrogenase (complex I) from the strictly aerobic yeast Yarrowia lipolytica. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1459(1), 230-238.

Flynn, P. J., & Reece, R. J. (1999). Activation of transcription by metabolic intermediates of the pyrimidine biosynthetic pathway. *MOLECULAR AND CELLULAR BIOLOGY*, 19(1), 882-888.

Garofano, A., Zwicker, K., Kerscher, S., Okun, P., & Brandt, U. (2003). Two Aspartic Acid Residues in the PSST-Homologous NUKM Subunit of Complex I from Yarrowia lipolytica Are Essential for Catalytic Activity. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 278(43), 42435-42440.

Grgic, L., Zwicker, K., Kashani-Poor, N., Kerscher, S., & Brandt, U. (2004). Functional Significance of Conserved Histidines and Arginines in the 49-kDa Subunit of Mitochondrial Complex I. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 279(20), 21193-21199.

Grigorieff, N. (1998). Three-dimensional structure of bovine NADH:ubiquinone oxidoreductase (complex I) at 22 A in ice. *JOURNAL OF MOLECULAR BIOLOGY*, 277(5), 1033-1046.

Hameedi, M. A., Grba, D. N., Richardson, K. H., Jones, A. J., Song, W., Roessler, M. M. & Hirst, J. (2021). A conserved arginine residue is critical for stabilizing the N2 FeS cluster in mitochondrial complex I. *JOURNAL OF MOLECULAR BIOLOGY*, 100474.

Hinchliffe, P., Carroll, J., & Sazanov, L. A. (2006). Identification of a Novel Subunit of Respiratory Complex I fromThermus thermophilus†. *BIOCHEMISTRY*, 45(14), 4413-4420.

Hirst, J., & Roessler, M. M. (2016). Energy conversion, redox catalysis and generation of reactive oxygen species by respiratory complex I. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1857(7), 872-883.

Ingledew, W. J., & Ohnishi, T. (1980). An analysis of some thermodynamic properties of iron-sulphur centres in site I of mitochondria. *BIOCHEMICAL JOURNAL*, 186(1), 111-117.

Ingledew, W. J., & Poole, R. K. (1984). The respiratory chains of Escherichia coli. *Microbiol Rev*, 48(3), 222-271.

Kashani-Poor, N., Kerscher, S., Zickermann, V., & Brandt, U. (2001). Efficient large scale purification of his-tagged proton translocating NADH:ubiquinone oxidoreductase (complex I) from the strictly aerobic yeast Yarrowia lipolytica. *Biochim Biophys Acta*, 1504(2-3), 363-370.

Kashani-Poor, N., Zwicker, K., Kerscher, S., & Brandt, U. (2001). A Central Functional Role for the 49-kDa Subunit within the Catalytic Core of Mitochondrial Complex I. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 276(26), 24082-24087.

Kerscher, S., Dröse, S., Zwicker, K., Zickermann, V., & Brandt, U. (2002). Yarrowia lipolytica, a yeast genetic system to study mitochondrial complex I. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1555(1), 83-91.

Kerscher, S., Kashani-Poor, N., Zwicker, K., Zickermann, V., & Brandt, U. (2001). Exploring the catalytic core of complex I by Yarrowia lipolytica yeast genetics. *JOURNAL OF BIOENERGETICS AND BIOMEMBRANES*, 33(3), 187-196.

Le Breton, N., Wright, J. J., Jones, A. J. Y., Salvadori, E., Bridges, H. R., Hirst, J., & Roessler, M. M. (2017). Using Hyperfine Electron Paramagnetic Resonance Spectroscopy to Define the Proton-Coupled Electron Transfer Reaction at Fe–S Cluster N2 in Respiratory Complex I. *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, 139(45), 16319-16326.

Leif, H., Sled, V. D., Ohnishi, T., Weiss, H., & Friedrich, T. (1995). Isolation and characterization of the proton-translocating NADH: ubiquinone oxidoreductase from Escherichia coli. *Eur J Biochem*, 230(2), 538-548.

Meinhardt, S. W., Wang, D. C., Hon-nami, K., Yagi, T., Oshima, T., & Ohnishi, T. (1990). Studies on the NADH-menaquinone oxidoreductase segment of the respiratory chain in Thermus thermophilus HB-8. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 265(3), 1360-1368.

Mooser, D., Maneg, O., MacMillan, F., Malatesta, F., Soulimane, T., & Ludwig, B. (2006). The menaquinol-oxidizing cytochrome bc complex from Thermus thermophilus: Protein domains and subunits. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1757(9-10), 1084-1095.

Ohnishi, T. (1998). Iron-sulfur clusters/semiquinones in complex I. *Biochim Biophys Acta*, 1364(2), 186-206.

Reda, T., Barker, C. D., & Hirst, J. (2008). Reduction of the Iron–Sulfur Clusters in Mitochondrial NADH:Ubiquinone Oxidoreductase (Complex I) by EuII -DTPA, a Very Low Potential Reductant†. *BIOCHEMISTRY*, 47(34), 8885-8893.

Rhein, V. F., Carroll, J., Ding, S., Fearnley, I. M., & Walker, J. E. (2013). NDUFAF7 methylates arginine 85 in the NDUFS2 subunit of human complex I. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 288(46), 33016-33026.

Roessler, M. M., Evans, R. M., Davies, R. A., Harmer, J., & Armstrong, F. A. (2012). EPR spectroscopic studies of the Fe-S clusters in the O2-tolerant [NiFe]-hydrogenase Hyd-1 from Escherichia coli and characterization of the unique [4Fe-3S] cluster by HYSCORE. *JOURNAL OF THE AMERICAN CHEMICAL SOCIETY*, 134(37), 15581-15594.

Sazanov, L. A. (2015). A giant molecular proton pump: structure and mechanism of respiratory complex I. *Nat Rev Mol Cell Biol*, 16(6), 375-388.

Sharma, V., Belevich, G., Gamiz-Hernandez, A. P., Róg, T., Vattulainen, I., Verkhovskaya, M. L., Wikström, M., Hummer, G., & Kaila, V. R. I. (2015). Redox-induced activation of the proton pump in the respiratory complex I. *Proceedings of the National Academy of Sciences*, 112(37), 11571-11576.

Stoll, S., & Schweiger, A. (2006). EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *JOURNAL OF MAGNETIC RESONANCE*, 178(1), 42-55.

Tocilescu, M. A., Fendel, U., Zwicker, K., Dröse, S., Kerscher, S., & Brandt, U. (2010). The role of a conserved tyrosine in the 49-kDa subunit of complex I for ubiquinone binding and reduction. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1797(6-7), 625-632.

Wang, J. H., Hung, W., & Tsai, S. H. (2011). High efficiency transformation by electroporation of Yarrowia lipolytica. *JOURNAL OF MICROBIOLOGY*, 49(3), 469-472.

Wright, J. J., Salvadori, E., Bridges, H. R., Hirst, J., & Roessler, M. M. (2016). Small-volume potentiometric titrations: EPR investigations of Fe-S cluster N2 in mitochondrial complex I. *JOURNAL OF INORGANIC BIOCHEMISTRY*, 162, 201-206.

Yakovlev, G., Reda, T., & Hirst, J. (2007). Reevaluating the Relationship between EPR Spectra and Enzyme Structure for the Iron-Sulfur Clusters in NADH:Quinone Oxidoreductase. *Proceedings of the National Academy of Sciences - PNAS*, 104(31), 12720-12725.

Yang, Y., & Bedford, M. T. (2013). Protein arginine methyltransferases and cancer. *NATURE REVIEWS CANCER*, 13(1), 37-50.

Zhu, J., Vinothkumar, K. R., & Hirst, J. (2016). Structure of mammalian respiratory complex I. *NATURE*, 536(7616), 354-358.

Zickermann, V., Zwicker, K., Tocilescu, M. A., Kerscher, S., & Brandt, U. (2007). Characterization of a subcomplex of mitochondrial NADH:ubiquinone oxidoreductase (complex I) lacking the flavoprotein part of the N-module. *Biochim Biophys Acta*, 1767(5), 393-400.

Zwicker, K., Galkin, A., Dröse, S., Grgic, L., Kerscher, S., & Brandt, U. (2006). The Redox-Bohr Group Associated with Iron-Sulfur Cluster N2 of Complex I. *JOURNAL OF BIOLOGICAL CHEMISTRY*, 281(32), 23013-23017.

Appendix VIII

The primers used in Appendix VII

Table B6 Primers used for making and sequencing the *NUCM* point mutation

Primer Name	Sequence (5'-3')
R121LYS-F	CTGCACAAGGGAACCGAGAAGC
R121LYS-R	CAGACCGACGTGGGGGTC
49-2.5	GAGACCTACACCGCCATCGAG
n49knc1	ATGTGTGGAAGTCTGACC
17899	TCACACAATGCTGCGATC
NUCM-R	AGCAACTCTGCTCTCACATACA

Table B7 Primers used for making and sequencing the *YALI0D12859g*-related constructs

Primer Name	Sequence (5'-3')
URA3-NcoI-F	CAACGCTGGCCATGGGTCGACAAAGGCCTGTTTCT
URA3-EcoRV-R	ATCCCACGGGATATCGTCGACGAGTATCTGTCTGA
URA-seq-F	CTATTGAGGAGGCCAAGCGATAC
URA-seq-R	GACTTGTGGACGTTAGCTCGAGC
12859-homo-F	ggtacccggggatccCTTCCCGTCTCCATCCTTGT
12859-homo-R	gactctagaggatccAGTTCATCGATCTGTTTGTG
12859-seq-F	TGGCGGCGTAGATAATAGCTC
12859-seq-R	TATTGACCGACAACCACCGCG